UNIVERSITY OF NATIONAL AND WORLD ECONOMY

The Spirit Makes The Power

15TH INTERNATIONAL CONFERENCE ON APPLICATION OF INFORMATION AND COMMUNICATION TECHNOLOGY AND STATISTICS IN ECONOMY AND EDUCATION

ICAICTSEE – 2025

October 9 - 11, 2025
University of National and World Economy
Sofia, Bulgaria

CONFERENCE PROCEEDINGS

http://icaictsee.unwe.bg/

UNDER THE AUSPICES OF:

ifip
the leading edge of information technology

PUBLICATIONS INDEXED BY:

CEEOL

COOPERATED BY

CREATIVITY AND HARMONY SOCIETY

PROCEEDINGS OF THE

15TH INTERNATIONAL CONFERENCE ON APPLICATION OF INFORMATION AND COMMUNICATION TECHNOLOGY AND STATISTICS IN ECONOMY AND EDUCATION

October 9 - 11, 2025, Sofia Bulgaria

ICAICTSEE-2025

http://icaictsee.unwe.bg/

EDITORS:

PROF. DR. DIMITER G. VELEV
DEPARTMENT OF INFORMATICS
UNIVERSITY OF NATIONAL AND WORLD ECONOMY, SOFIA, BULGARIA
DGVELEV@UNWE.BG

PROF. DR. SERGHEI OHRIMENCO
LABORATORY OF INFORMATION SECURITY
ACADEMY OF ECONOMIC STUDIES OF MOLDOVA, CHISINAU, MOLDOVA

PROF. DR. PLAMENA ZLATEVA
DEPARTMENT OF INFORMATICS
UNIVERSITY OF NATIONAL AND WORLD ECONOMY, SOFIA, BULGARIA

PROF. DR. SIVAKUMAR R.
SCHOOL OF ELECTRONICS ENGINEERING
VELLORE IINSTITUTE OF TECHNOLOGY, VELLORE, TAMIL NADU, INDIA

ISSUED FOR PUBLICATION: APRIL 20th, 2026

ISSN: 2367-7635 (PRINT)

ISSN: 2367-7643 (ONLINE)

ACADEMIC PUBLISHING HOUSE OF UNWE, SOFIA, BULGARIA

Conference Committees

CONFERENCE CHAIR

Dimiter G. **Velev**	University of National and World Economy	**Bulgaria**

PROGRAM COMMITTEE

César Correa **Arias**	Universidad de Guadalajara	**Mexico**
Saji **Baby**	GEO Environmental Consultation	**Kuwait**
Andrei **Buiga**	Artifex University of Bucharest	**Romania**
Marian-Cristian **Barbu**	Artifex University of Bucharest	**Romania**
Gaetano **Cascini**	Politecnico di Milano	**Italy**
Luisa **Chimenz**	University of Genoa	**Italy**
Ghenadie **Ciobanu**	Artifex University of Bucharest	**Romania**
Sérgio Manuel Serra da **Cruz**	Universidade Federal Rural do Rio de Janeiro	**Brazil**
Tadeusz **Czachorski**	Institue of Theoretical and Applied Informatics, PAS	**Poland**
Linda **Daniela**	University of Latvia	**Latvia**
Vladimir **Dimitrov**	Sofia State University	**Bulgaria**
Marco **Ferrante**	University of Genoa	**Italy**
Carlos Arturo T. **Gastelu**	Universidad Veracruzana	**Mexico**
Chih-Hsuan **Huang**	Hubei University of Economics	**China**
Aytekin **Isman**	Sakarya University	**Turkey**
Hee Dong **Kim**	Hankuk University of Foreign Studies	**S. Korea**
Rosen **Kirilov**	University of National and World Economy	**Bulgaria**
Alexandru-Lucian **Manole**	Artifex University of Bucharest	**Romania**
Kaninda **Musumbu**	Universite Bordeaux	**France**
Tetsuo **Noda**	Shimane University	**Japan**
Serghei **Ohrimenco**	Laboratory of Information Security, AESM	**Moldova**
David Nadler **Prata**	Universidade Federal do Tocantins	**Brazil**

N. **Sharmila**	Vellore Institute of Technology	**India**
R. **Sivacoumar**	Vellore Institute of Technology	**India**
Rajagopal **Sivakumar**	Vellore Institute of Technology	**India**
Tullio **Tanzi**	Institute Mines Telecom - Telecom ParisTech	**France**
S**. Vidhya**	Vellore Institute of Technology	**India**
Hsin-Hung **Wu**	National Changhua University of Education	**Taiwan**
Teodora **Zapryanova**	University of Economics - Varna	**Bulgaria**
Plamena **Zlateva**	University of National and World Economy	**Bulgaria**
Zong **Xuejun**	Shenyang University of Chemical Technology	**China**

UNWE ORGANIZING COMMITTEE

Katya **Kirilova**	University of National and World Economy	**Bulgaria**
Plamen **Milev**	University of National and World Economy	**Bulgaria**
Venko **Andonov**	University of National and World Economy	**Bulgaria**
Gabriela **Naskova**	University of National and World Economy	**Bulgaria**
Desislava **Stoyanova**	University of National and World Economy	**Bulgaria**

Dear Colleagues,

*On behalf of the Conference Committees I have the honor and pleasure to thank you all for your true participation in the work of the **15th International Conference on Application of Information and Communication Technology and Statistics in Economy and Education (ICAICTSEE-2025)**, http://icaictsee.unwe.bg/, which took place on October 9 - 11, 2025 at the University of National and World Economy (UNWE), Sofia, Bulgaria.*

I would like specially to express my gratitude to all of the program committee members for their genuine support without which it would never had happened.

Due to the current state of scientific development in all spheres of human activity, the constant knowledge and skill actualization of the academia and researchers in the field of Information and Communication Technology (ICT) is an obligatory necessity, especially when the world is in a global economic and financial crisis. The definition of long-term scientific research tasks in this area has a strategic importance, which is even truer for young researchers, lecturers and PhD. students.

The conference aims were conducting analysis of the current problems and presenting results of the ICT application in different areas of economy, education and related areas of scientific knowledge; outlining the existing possibilities for the application of modern ICT tools, methods, methodologies and information systems in economy and education; discussing advanced and emerging research trends with a long-term importance in the field of ICT application in economy and education.

The conference has established an academic environment that fosters the dialogue and exchange of ideas between different levels of academic and research communities.

Dimiter G. Velev
ICAICTSEE-2025 Chair

CONTENTS

TITLE	Pages
ONTOLOGY OF SOFTWARE ABUSE **Serghei Ohrimenco[1], Ghenadie Belinschi[1], Liudmyla Rybalchenko[2], Eduard Ryzhkov[3]** [1] Academy of Economic Studies of Moldova, Chisinau, Moldova [2] University of Customs and Finance, Dnipro, Ukraine [3] Dnipro State University of Internal Affairs, Dnipro, Ukraine	**1**
DIGITAL TRANSFORMATION OF MOLDOVA'S AGRI-FOOD SECTOR: RISKS OF CREATING E-COMMERCE PLATFORMS FOR FARM PRODUCTS CRM **Aurelia Turcan** Academy of Economic Studies of Moldova, Chisinau, Moldova	**10**
E-COMMERCE: INNOVATIVE TOOLS FOR DIGITALIZING TRADE IN MOLDOVA **Aurelia Turcan, Lucia Gujuman** Academy of Economic Studies of Moldova, Chisinau, Moldova	**22**
USING ARTIFICIAL INTELLIGENCE TO ENSURE INFORMATION SECURITY OF ORGANIZATIONS: POSSIBILITIES, PROBLEMS AND PROSPECTS **Olga V. Pugacheva** Francisk Skorina Gomel State University, Gomel, Belarus	**35**
NVIVO APPLICATION FOR QUALITATIVE DATA ANALYSIS IN SCIENTIFIC RESEARCH: OPTIONS AND RESULT VISUALIZATIONS **Tatiana Colesnicova, Mihail Ciobanu** Academy of Economic Studies of Moldova, Chisinau, Moldova	**46**
CONSTRUCTING A CONCEPTUAL MODEL FOR THE MODERNIZATION OF A SMALL FAMILY-OWNED ENTERPRISE **Ivan Stamov** Academy of Economic Studies of Moldova, Chisinau, Republic of Moldova	**57**
INFORMATION AND TELECOMMUNICATION TECHNOLOGIES AS THE BASIS OF MODERN CYBER SECURITY **Liudmyla Rybalchenko** University of Customs and Finance, Dnipro, Ukraine	**69**
GLOBAL TRENDS AND NATIONAL REALITIES IN ESG REPORTING **Irina Iavorschi, Liliana Lazari** Academy of Economic Studies of Moldova, Chisinau, Moldova	**73**
STOCHASTIC MODELING OF CYBER RESILIENCE IN SMART POWER GRIDS USING CONTINUOUS-TIME MARKOV CHAINS **Svetlana Konyuk, Alexey Rusanov** S.C. "Add-Production" S.R.L., Chisinau, Moldova	**85**

INVESTMENTS IN MOLDOVA'S ENERGY INFRASTRUCTURE: PRIORITIES AND MODERNIZATION PROSPECTS **Timur Lesan** Academy of Economic Studies of Moldova, Chisinau, Moldova	**93**
MIND GAMES AND MACHINES: SOCIAL ENGINEERING IN THE CYBERSECURITY ERA **Bradu Nichita** IU International University of Applied Sciences, Berlin, Germany	**100**
WHY WE NEED NEW ACCESS MODELS: FROM SESSION-BOUND AUTHORIZATION TO CONTINUOUS ACCESS **Ghenadie Belinschi** Academy of Economic Studies of Moldova, Chisinau, Moldova	**108**
GLOBAL TRENDS IN INTELLECTUAL PROPERTY IN 2023–2024 **Grisciuc-Bucica Svetlana** Academy of Economic Studies of Moldova, Chisinau, Moldova	**114**
A REVIEW ON VARIOUS DEEP LEARNING TECHNIQUES FOR THE DETECTION OF BRAIN DISORDERS **Enteamen Yengkhom, Sivakumar Rajgopal** Vellore Institute of Technology, Vellore, Tamil Nadu, India	**124**
MACHINE LEARNING AND DEEP LEARNING FOR CUFF-LESS BLOOD PRESSURE MONITORING: A REVIEW OF ADVANCES AND FUTURE DIRECTIONS **Sneha Majumdar, Sivakumar Rajgopal** Vellore Institute of Technology, Vellore, Tamil Nadu, India	**135**
MACHINE LEARNING AND DEEP LEARNING IN AUDIOLOGY: PATHWAYS TO IMPROVED DIAGNOSIS AND ACCESS **T Priya Dharsini, Sivakumar Rajgopal** Vellore Institute of Technology, Vellore, Tamil Nadu, India	**144**
SYNTHETIC PPG: DATA GENERATION AND APPLICATIONS **Diptman Gupta, Sivakumar Rajagopal** Vellore Institute of Technology, Vellore, Tamil Nadu, India	**151**
MACHINE LEARNING IN HEALTHCARE STAFF TRAINING AND ASSESSMENT **R Rishab, Sivakumar Rajagopal** Vellore Institute of Technology,Vellore, Tamil Nadu, India	**160**
RECENT TRENDS IN ORGANIZATIONAL STRUCTURES: TOWARDS AGILE AND DIGITALLY-ENABLED ENTERPRISES **Ayushmaan Panda, Sivakumar R** Vellore Institute of Technology, Vellore, Tamil Nadu, India	**169**
RECENT ADVANCES IN HOSPITAL INFORMATION SYSTEMS: TOWARDS SMARTER HEALTHCARE **Prarthana M, Sivakumar R** Vellore Institute of Technology, Vellore, Tamil Nadu, India	**176**

ICU BED PREDICTION MODEL USING ENSEMBLE LEARNING **Adhvaith Shivkumar, Sivakumar Rajagopal** Vellore Institute of Technology, Vellore, Tamil Nadu, India	**188**
A REVIEW OF DEEP LEARNING APPROACHES IN POSTPARTUM DEPRESSION DETECTION **Vivek Saha, Nehal Chhalotre, Aryan Chaurasia, Sivakumar R** Vellore Institute of Technology, Vellore, Tamil Nadu, India	**195**
RECENT TRENDS IN SOFTWARE DEVELOPMENT LIFE CYCLE (SDLC) IN HEALTHCARE **Siva Renjini Jeevan, Sivakumar Rajgopal** Vellore Institute of Technology, Vellore, Tamil Nadu, India	**205**
A SYSTEMATIC REVIEW ON ADVANCED MULTI-CHANNEL AND WEARABLE SURFACE ELECTROMYOGRAPHY SYSTEMS WITH AI-DRIVEN SIGNAL PROCESSING AND APPLICATIONS IN CLINICAL DIAGNOSTICS AND PROSTHETIC CONTROL **Triparna Maitra, Sivakumar R** Vellore Institute of Technology, Vellore, Tamil Nadu, India	**208**
TIME SERIES-BASED STAFFING FORECASTING USING ARIMA AND SARIMA MODELS FOR WORKFORCE OPTIMIZATION **Neeraj Srinivasan, Sivakumar Rajagopal** Vellore Institute of Technology, Vellore, Tamil Nadu, India	**216**
SYMPTOM-BASED DISEASE PREDICTION USING MACHINE LEARNING AND NLP **Vignesh.T.A, Sivakumar R** Vellore Institute of Technology, Vellore, Tamil Nadu, India	**224**
MACHINE LEARNING-BASED TRIAGE SYSTEM FOR EFFICIENT TELEMEDICINE CALL MANAGEMENT **Alphonse Vimal Jose J, Sivakumar R** Vellore Institute of Technology, Vellore, Tamil Nadu, India	**230**
MACHINE LEARNING APPLICATIONS IN BIOMEDICAL TELEMETRY SYSTEMS FOR DROWSINESS DETECTION: A REVIEW **Dhanush Kumar V, Sivakumar Rajgopal** Vellore Institute of Technology, Vellore, Tamil Nadu, India	**239**
EXPLORING THE POTENTIAL OF BACTERIA-MEDIATED CHROMIUM CHELATION IN BIOMEDICAL WASTE CONTAMINATED WATER **Kanthasami Vijaayganapathi, Saqib Hassan1, Kunal Biswas** Sathyabama Institute of Science and Technology, Tamil nadu, Chennai, India	**245**
THE DIGITAL TRANSFORMATION OF TALENT ACQUISITION: A REVIEW OF AI-DRIVEN STRATEGIES AND ETHICAL CONSIDERATIONS **Gokul Nittin A, Sivakumar Rajagopal** Vellore Institute of Technology, Vellore, Tamil Nadu, India	**256**

INFORMATION SECURITY AUDIT – AN OBJECTIVE PREREQUISITE FOR IMPROVING THE CYBERSECURITY OF THE ENTERPRISE **Rositsa Ivanova** University of National and World Economy, Sofia, Bulgaria	**263**
ECONOMIC METRICS FOR MEASURING THE EFFECTIVENESS AND EFFICIENCY OF CYBERSECURITY INVESTMENTS **Nedko Tagarev** University of National and World Economy, Sofia, Bulgaria	**274**
UTILIZING ARTIFICIAL INTELLIGENCE IN AUDITING AND INVESTIGATIONS **Iavor Bachev** University of National and World Economy, Sofia, Bulgaria	**283**
CYBERSECURITY RISK MANAGEMENT AND THREAT-LED PENETRATION TESTING FOR FINANCIAL STABILITY **Maya Zhelyazkova** University of National and World Economy, Sofia, Bulgaria	**287**
COUNTERING GENERATIONAL DISENFRANCHISEMENT: BIPARTISAN STRATEGIES TO ENHANCE HOMEOWNERSHIP FOR MILLENNIALS AND GENERATION Z **Boyan Mitrakiev** University of National and World Economy, Sofia, Bulgaria	**298**
ARTIFICIAL INTELLIGENCE IN MANAGING FORCED MIGRATION: EVIDENCE FROM EUROPE **Vasil Pavlov** University of National and World Economy, Sofia, Bulgaria	**309**
REGIONAL FOCUS AND SENTIMENT TRENDS IN BULGARIAN MEDIA: A STUDY BASED ON WEB PUBLICATIONS **Plamen Milev** University of National and World Economy, Sofia, Bulgaria	**315**
METHODOLOGICAL DIMENSIONS IN STRUCTURING CONTENT ON VIDEO PLATFORMS **Yavor Tabov** University of National and World Economy, Sofia, Bulgaria	**322**
A TAXONOMY OF SIMULATION - BASED EVALUATION FOR ONLINE MACHINE LEARNING **Bozhidar Bahov** University of National and World Economy, Sofia, Bulgaria	**329**
SOCIAL SECURITY – CONTEMPORARY ASPECTS (THE EDUCATION) **Konstantin Poudin** University of National and World Economy, Sofia, Bulgaria	**342**

AI FOR DETECTING AND COMBATING DEEPFAKES AND DISINFORMATION **Veronica Hîncu** Academy of Economic Studies of Moldova, Chisinau, Moldova	**350**
SOME WAYS TO DEVELOP A CULTURE OF CYBER SECURITY AMONG TEACHING STAFF **Liubimir Chiriac, Violeta Bogdanova** "Ion Creangă" State Pedagogical University, Chisinau, Moldova	**357**
ARTIFICIAL INTELLIGENCE IN THE THEORY AND PRACTICE TO CONSUMER BEHAVIOR **Irina Balina** Academy of Economic Studies of Moldova, Chisinau, Moldova	**363**
COMPARATIVE ANALYSIS OF INTERNATIONAL DIGITALIZATION MODELS FOR AGRITOURISM DEVELOPMENT **Svetlana Ceban** Academy of Economic Studies of Moldova, Chisinau, Moldova	**371**
LISTENING IN ENGLISH FOR IT STUDENTS: CHALLENGES AND SOLUTIONS **Anastasia Romanova** Academy of Economic Studies of Moldova, Chisinau, Moldova	**380**
DECISION SUPPORT SYSTEMS IN OPERATIONAL RISK MANAGEMENT: TRENDS, PERSPECTIVES AND THREATS **Olga Caminschi** Academy of Economic Studies of Moldova, Chisinau, Moldova	**386**
THE CIRCULAR SUPPLY CHAIN MODEL IN REDUCING FOOD WASTE AND REUSING RESOURCES IN THE FOOD INDUSTRY **Eugeniu Stoian** Academy of Economic Studies of Moldova, Chisinau, Moldova	**395**
MEASURING KNOWLEDGE-BASED COMPETITIVENESS THROUGH ICT AND INNOVATION INDICATORS: A MULTIDIMENSIONAL APPROACH **Diana Ignatiuc** Academy of Economic Studies of Moldova, Chisinau, Moldova	**405**
ECONOMIC VALUE OF AI-DRIVEN INDOOR AIR QUALITY MONITORING IN SMART HOMES: A COST-BENEFIT ANALYSIS **Adriana Catruc** Academy of Economic Studies of Moldova, Chisinau, Moldova	**421**
AN ANALYTICAL REVIEW OF THE LITERATURE ON THE IMPACT OF TRAFFIC FLOW AND WEATHER CONDITIONS ON ROAD SAFETY **Silvia Assenova** Institute of Robotics, Bulgarian Academy of Sciences, Sofia, Bulgaria	**426**

DATA MODELS IN NETWORK-CENTRIC SYSTEMS FOR DISASTER RISK MANAGEMENT **Georgi Kirov** International Business School, Botevgrad, Bulgaria	**438**
THE INSTITUTIONALIZATION OF KOREAN STUDIES IN BULGARIA: FROM FOUNDATIONAL DRIVERS TO FUTURE PROSPECTS **Hwan Joung James Kang** University of Plovdiv "Paisii Hilendarski", University of Veliko Tarnovo "St. Cyril and St. Methodius," Bulgaria	**448**
PRINCIPLES OF DEVELOPING SECURE SOFTWARE APPLICATIONS **Rosen Kirilov** University of National and World Economy, Sofia, Bulgaria	**458**
ASSESSMENT OF THE RISK MANAGEMENT SYSTEM OF PENSION INSURANCE COMPANIES **Maia Natchkova** University of National and World Economy, Sofia, Bulgaria	**462**
USING MATHEMATICAL METHODS TO SOLVE ECONOMIC PROBLEMS AND PRODUCTION CRISES IN ENERGY AND TOURISM **Kostadin Sheiretsky, Svetlin Antonov** Technical University of Sofia	**474**
WILDFIRES AS ECONOMIC DISRUPTORS: SUSTAINABLE POLICIES FOR ENERGY, TOURISM AND REGIONAL RECOVERY **Kostadin Sheiretsky, Svetlin Antonov** Technical University of Sofia	**479**
DATA AS A GOVERNANCE TOOL: BIG DATA AND SME EXPORT PROMOTION IN SOUTH KOREA **Hannah Kang** University of National and World Economy, Sofia, Bulgaria	**483**
TESTING THE DATA ACCESS LAYER: MODERN APPROACHES AND AI ASSISTANCE **Maria Marzovanova, Veska Mihova** University of National and World Economy, Sofia, Bulgaria	**494**
METADATA-BASED UNIT TESTING OF SQL SERVER STORED PROCEDURES WITH TSQLT **Veska Mihova, Maria Marzovanova** University of National and World Economy, Sofia, Bulgaria	**501**
AI AIDING FINANCIAL CONTROL **Milen Markov** University of National and World Economy, Sofia, Bulgaria	**508**

A MODEL-DRIVEN APPROACH TO INFORMATION SECURITY OF AUTOMATED AND AUTONOMOUS SYSTEMS IN INDUSTRY 4.0 ENVIRONMENTS **Ivan Gaidarski** Institute of Robotics, Bulgarian Academy of Sciences, Sofia, Bulgaria	**515**
INSIDER THREATS IN AUTOMATED AND AUTONOMOUS SYSTEM ENVIRONMENTS **Ivan Gaidarski** Institute of Robotics, Bulgarian Academy of Sciences, Sofia, Bulgaria	**525**
OPPORTUNITIES FOR USING DIGITAL TWINS IN NATURAL DISASTER RISK MANAGEMENT **Hristina Angelova** University of National and World Economy, Sofia, Bulgaria	**535**
SPECIFIC FEATURES OF ENTERPRISE LIQUIDITY ANALYSIS **Bojidar Nachev** University of National and World Economy, Sofia, Bulgaria	**541**
CHALLENGES TO THE NATURAL DISASTER PROTECTION SYSTEM IN THE REPUBLIC OF BULGARIA **Stefan Stefanov, Ekaterina Bogomilova** University of National and World Economy, Sofia, Bulgaria	**546**
SOME ASPECTS OF THE CONNECTION BETWEEN ARTIFICIAL INTELLIGENCE, SUSTAINABLE BUSINESS AND CLIMATE CHANGE **Milena Kunova, Katerina Doychinova** University of National and World Economy, Sofia, Bulgaria	**558**
SELF-SERVICE SYSTEMS IN THE PRIVATE AND PUBLIC SECTOR: ADOPTION, BENEFITS, AND CHALLENGES **Petar Genov** University of National and World Economy, Sofia, Bulgaria	**564**
SECURITY AND PRIVACY CONCERNS IN SELF-SERVICE SYSTEMS **Petar Genov** University of National and World Economy, Sofia, Bulgaria	**572**
OPPORTUNITIES FOR INCREASING THE POTENTIAL OF THE BULGARIAN DEFENSE INTELLIGENCE SERVICE THROUGH THE VOLUNTARY ACTIVE RESERVE OF THE BULGARIAN ARMED FORCES **Stanimir Stoyanov** University of National and World Economy, Sofia, Bulgaria	**577**
CORRUPTION RISK ASSESSMENT IN PUBLIC PROCUREMENT **Stela Stanislavova** University of National and World Economy, Sofia, Bulgaria	**584**

AN ANALYTICAL MODEL FOR ASSESSING CORRUPTION RISK IN PUBLIC PROCUREMENT: EVIDENCE FROM SENSITIVE SECTORS **Stela Stanislavova** University of National and World Economy, Sofia, Bulgaria	**591**
KEY CORRUPTION RISKS IN BULGARIAN PUBLIC PROCUREMENT **Stela Stanislavova** University of National and World Economy, Sofia, Bulgaria	**600**
SELECTION OF AN ALTERNATIVE FOR REDUCING THE RISK OF FOREST FIRES IN THE BLACK SEA REGION THROUGH THE APPLICATION OF A PAYOFF MATRIX AND SIMULATION MODELING **Anna-Maria Markova, Yasen Tsvetkov** University of National and World Economy, Sofia, Bulgaria	**607**
TECHNOLOGICAL INNOVATIONS AND COOPERATIVE FRAMEWORK IN NUCLEAR SECURITY: STRENGTHENING INTEGRATED BORDER SECURITY SYSTEM FOR GLOBAL STABILITY **Sara Khalid Khider Mohamed** University of National and World Economy, Sofia, Bulgaria	**613**
EDUCATION AND TRAINING IN NUCLEAR SECURITY IN SUDAN – STATUS AND PROSPECTS **Bokhari Ahmed Moneir Ahmed** University of National and World Economy, Sofia, Bulgaria	**626**
APPLICATION OF MACHINE LEARNING IN THE FINANCIAL SECTOR **Kameliya Petkova and Melani Lyubenova** University of National and World Economy, Sofia, Bulgaria	**637**

Ontology of Software Abuse

Serghei Ohrimenco[1], Ghenadie Belinschi[1], Liudmyla Rybalchenko[2], Eduard Ryzhkov[3]

[1] Information Security Lab., Academy of Economic Studies of Moldova, Chisinau, Moldova
[2] University of Customs and Finance, Dnipro, Ukraine
[3] Dnipro State University of Internal Affairs, Dnipro, Ukraine
osa@ase.md, ghenadie.belinschi@ase.md, luda_r@ukr.net, revord924@ukr.net

Abstract. The problem of classification and identification of software abuse does not lose its relevance. The aim of this paper is to present the evolution of software abuse based on historical and technological context and ontology formation. A review of the works of leading experts in the following areas is presented: computer viruses, computer attacks, threats to mobile devices and AI for Science. The authors analyses the composition of malicious programs, identifying such groups as Virus, Worm, Trojan, Botnet, Ransomware and others. Special attention is paid to the study of stages in the process of malware evolution and the formation of a set of service models such as Cybercrime-as-a-Service, Ransomware-as-a-Service and others. It is noted that the development of malware is directly related to the formation and functioning of the shadow digital economy.

Keywords. Soft Abuse, Computer Virus, Ontology, Malware, Vulnerabilities, Worm, Attack

1. Introduction

To define corporate sustainability is not an easy task. Even till now terms like CSR (Corporate social responsibility) and corporate citizenship are continue widely to be used. Nowadays, however, all these concepts are replaced by the broader term corporate sustainability. In the beginning it was not customary to define sustainability at corporate level.

The purpose of this report is to review the most common concepts of corporate sustainability in order to outline basic qualitative aspects of the most popular views. The ultimate goal is to find promising directions for building quantitative models of corporate sustainability at a later stage.

The most significant consideration is that these quantitative models had to be built on the basis of information available to the organization. This is information that is collected on a daily basis in corporate databases as a result of current activities.

2. Problem Statement

The authors set the goal of analyzing the emergence and development of software abuse (SA). This study covers a long period of development, taking into account their organization, the structure of their logical mechanisms, the purposes of their implementation and distribution, and methods of countering them.

Our research does not begin with an attempt to establish the authorship of SA, but rather by identifying the main sources of information. Unfortunately, history has not preserved the names of the developers (with rare exceptions – the Morris worm and the Brain computer

virus) or the specific conditions under which SA ideas were implemented, written, tested, and distributed. It should be emphasized that the practice of studying specific SAs shows that analysts and virologists are only confronted with the results of the implementation of internal mechanisms, that is, the final phase of the life cycle. All previous stages of the life cycle (including planning, requirements analysis, design, development, testing, implementation, and maintenance) remain completely inaccessible to researchers. Therefore, the main task is to create conditions for the formation of a software ontology – a formalized knowledge system that describes the methods of malicious use of software (standard and modified or specialized) in order to cause damage to the user or the information system.

During the preparation of this material, a debate arose regarding the use of two terms – ontology [1] and taxonomy [2], as they are frequently used in publications devoted to information technology. The choice fell on ontology, as a system for formally describing a domain of knowledge, categorizing viruses, their behavior, and relationships, as opposed to a simple hierarchy – taxonomy.

This material is based on the historical and technological context that characterizes the evolution of software – from the boot virus to the creation of service platforms and a specialized ecosystem.

3. Related Work

Most cybersecurity experts agree on the origins of computer viruses. Fred Cohen and his dissertation "Computer Viruses" (1986), as well as his earlier works [3], [4], [5], are considered the most influential.

A vast number of works have been devoted to the problems of describing the mechanisms of implementation and means of countering computer viruses. Among them, the leading authors, grouped into the following categories, should be highlighted:

1). Computer Viruses: Nikolai Bezrukov [6], Serghei Ohrimenco [7], Monnappa K. A. [8], P Fites, P Johnston and M Kratz [9], J. Parikka [10], P. Szor [11], J. Aycock [12], M. Cristodorescu [13], Sushi Jajodia [14], and others. It should be noted that the selection is difficult due to the large number of publications, and the list includes only the most representative ones. For example, books by N. Bezrukov and S. Ohrimenco were the first published in the USSR and largely determined the development of virology. Of particular note is the fundamental book by R. Burger [15], which provided the training and education for the first virologists and information security specialists.

2). Computer Attacks: Jakobsson, M., & Ramzan, Z. [16], Sotiropoulos, J. [17].

3). Threats to Mobile Devices: Misra, A., & Dubey, A. [18].

4). Recently, a new scientific and practical field has emerged – AI for Science, which utilizes artificial intelligence methods, machine learning models, and other components to create powerful security tools. These include: automated detection and response (SOAR/EDR), behavioral analysis, combating phishing and deepfakes, proactive threat hunting, etc. The main publications are the following: A. Hassan et al. [19], Khambam, S. K. R. et al. [20].

The work of renowned Bulgarian expert Dr. Veselin Bonchev [21], a pioneer of computer virology in Bulgaria and head of the first Computer Virology Laboratory at the Bulgarian Academy of Sciences, deserves special mention. His regular column and publications in Computers & Security magazine, which presented detailed characteristics of computer viruses, classifications of their destructive actions, and countermeasures, are also noteworthy. For that time (the mid- and late 1980s), this material was highly relevant and in demand.

We also consider it necessary to provide examples of ontologies. These include the following: Cybersecurity Ontology - The Relationship between Vulnerabilities, Standards, Legal, and Regulatory Requirements [22], A Milestone in Cybersecurity Ontology [23], Cyber Domain Ontology [24], The Design and Application of a Unified Ontology for Cyber Security [25], Cybersecurity Ontology for Dynamic Analysis of IT Systems [26], An Ontology-Based Cybersecurity Framework for AI-Enabled Systems and Applications [27].

An important aspect is a historical perspective on the field under study. January 2026 marked the 40th anniversary of the appearance of the "Amvi Brothers Virus" or Brain (also known as the Pakistani Brain or Lahore Brain). This was the first virus in history targeting the IBM PC. The virus infected the boot sector of five-inch floppy disks, slowing down the drive and occupying 7 kilobytes of RAM, making it inaccessible to the DOS system. The virus code included the real names, addresses, and phone numbers of the brothers who developed it, so that victims could contact them for "vaccination".

4. Methodology

The material in this article is based on an analysis of the available literature and a structural synthesis of data aimed at identifying and addressing problems related to the definition and classification of software abuse.

The methodological approach includes the following steps:

- formation of the research framework. This structure is derived from the specifics of the problem under study and a set of key questions, including characteristics and definitions of software abuses, their groups and types by purpose and triggering mechanisms;
- literature review. This study used the methodology of systematic historical review of scientific and practical literature, to synthesize existing knowledge and identify new trends related to the development of new groups of software abuses, whose activities are based on aggregation with topical components - artificial intelligence, machine learning, etc.;
- data analysis and synthesis. All collected data were systematized into thematic areas: historical stages of development of computer virology, principles of virus propagation, new types and mechanisms of implementation.

5. Results and Discussion

The four horsemen of the Apocalypse have been identified by the UN Secretary-General as the main threats to humanity: Geostrategic Tensions, Climate Change, Growing Mistrust at the Global level and the Danger of new Technologies. Speaking to delegates of the UN General Assembly, António Guterres outlined his priorities for 2020 [28]. In his report, he highlights a section entitled 'The "Flip Side" of the digital revolution', in which he points out that new technologies are developing so fast that we do not have time not only to react to them, but sometimes even to understand their essence. While they offered great benefits, digital technologies and artificial intelligence were becoming a tool for incitement, misinformation, interference with privacy, exploitation and crime. Thus, he considers unacceptable the use of autonomous combat systems - 'machines capable of killing without any human involvement and without any responsibility' - and calls for a ban on the use of 'Killer Robots'.

A.Guterres urged to restore order in cyberspace, which he called the digital 'Wild West'. "Terrorists, white supremacists and others like them abuse the Internet and social media," the UN chief said. - Bots spread misinformation, push polarisation and undermine democracy. Cybercrime will cost us $6 trillion next year".

Another front to combat the costs of new technologies is the labour market, where A.Guterres predicts that automation will put hundreds of millions of workers out of work over

the next decade. He sees a way out in reforming the education system to teach people to be lifelong learners. According to the Secretary-General, the UN is the most appropriate platform for governments, the private sector, and civil society to counter the digital divide with global co-operation.

Fifteen years before this speech, Aycock J. [12] had identified four key threats - the four horsemen of the electronic Apocalypse -. Spam, Bugs, Denials of Service, and Malicious Software, which are summarised in the following figure.

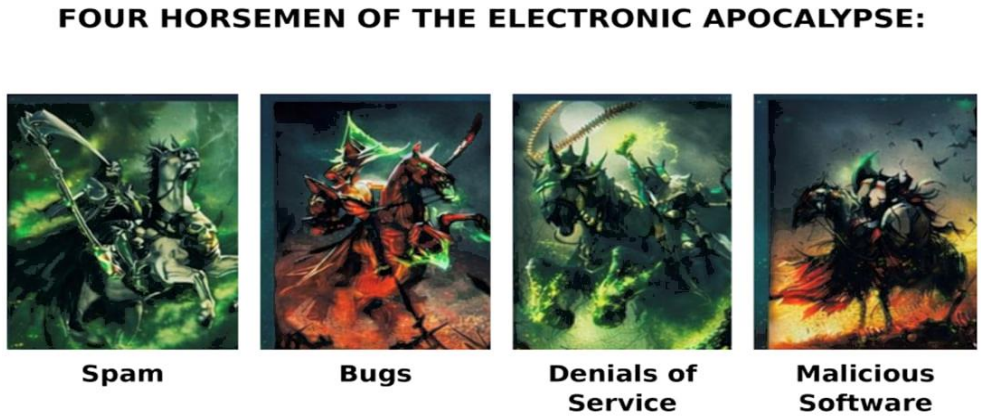


Fig.1 Four Horsemen of the Electronic Apocalypse, [©Siran Maxim 2025]

Malware is a broad term that refers to different types of malicious software such as viruses, worms, and rootkits. While performing malware analysis, experts come across different types of malware; some of them are classified based on their functionality and attack vectors:

- Virus or Worm: malware capable of copying itself and spreading to other computers. A virus requires user intervention, while a worm can spread without user intervention;
- Trojan: malware that masquerades as a regular programme to trick users into installing it on their systems. Once installed, it can perform malicious activities such as stealing sensitive data, uploading files to the attacker's server, or monitoring webcams;
- Backdoor/remote access Trojan (RAT): this is a type of Trojan that allows an attacker to access and execute commands on a compromised system;
- Adware: malware that displays unwanted adverts to the user. It is usually delivered via freeware downloads and can force software installation;
- Botnet: a group of computers infected with the same malware (called bots) waiting to receive instructions from a command and control server controlled by the attacker. The attacker can then send commands to these bots, which can perform malicious actions such as DDOS attacks or sending spam emails;
- information stealer: malware designed to steal sensitive data such as bank credentials or keystrokes from an infected system. Some examples of this malware include keyloggers, spyware, sniffers and form-grabbers;
- Ransomware virus: malware that holds a system to ransom by locking users out of their computer or by encrypting their files;

- Rootkit: malware that gives an attacker privileged access to an infected system, hides its presence or the presence of other software;

- Downloader or Dropper: malware designed to download or install additional malicious components.

Monappa K. A. points out [8] that classifying malware based on functionality is not always possible, because a single piece of malware may contain multiple functions that can fall into the many categories just mentioned. For example, malware may include a worm component that scans the network looking for vulnerable systems and may apply another malicious component such as a backdoor or ransomware virus when successfully exploited.

Malware can also be classified based on the attacker's motives. For example, if the malware is used to steal personal, commercial or proprietary information for profit, the malware may be classified as criminal software or commodity malware. If the malware targets a specific organisation or industry to steal information and gather intelligence for the purpose of espionage, then it may be classified as targeted or spyware [29].

Computer viruses are programs, just like a spreadsheet or a word processor. This distorted picture of computer viruses is the reason that the spectrum of responses to this topic range from sympathetic laughter to a knowing grin up to a panic over viral infection.

Factual explanations about these matters have generally been ignored up to now. There have been reports of "promiscuous disks," of worms which eat through the computer, and "hard viruses which destroy the ROM." To bring a little clarity to this confusion, this book is intended to be a guide in the world of computer viruses [15]. An interesting comparison between a biological virus and a computer virus is given.

J.Aycock in [12], proposes the following classification which includes: Logic Bomb, Trojan Horse, Back Door, Virus Worm, Rabbit, Spyware, Adware Hybrids, Droppers, and Blended Threats, Zombies.

Strictly speaking, the type of malware that replicates by attaching itself to other programs, but colloquially refers to malware in general. Usually contains a payload that performs unauthorized functions such as displaying messages, and deleting or modifying files [30].

The following table summarizes the results of a comparison of the two approaches - classical and modern - to describing threats.

Table 1. Comparison of approaches to describing threats

Criteria	**Classical approach**	**Modern approach**
Environment	Local executables, OS libraries	Cloud microservices, Docker containers/KBs
Implementation mechanisms	Disks (8", 5.25", 3.5"), physical access, email	Supply Chain Attacks (hacking through legitimate software updates)
Target	Malfunction, data corruption	Ransomware, hidden mining, AI model theft
Detection	Signature analysis, checksum verification	AI-powered behavioral analysis, Zero Trust architecture

Source: authors' elaboration

Note: The algorithms for implementing the SA remain logically identical to those described earlier, only the scale and technology stack change.

6. Mobile Threats

The evolution of mobile devices and threats have evolved almost in parallel. It should be noted that modern smartphones have a larger attack surface compared to conventional mobile phones. Attackers are mainly targeting Android and IPhone platforms [31].

Experts have adopted an approach that divides threats into three main types: device-level threats, network-level threats, and application-level threats.

Common mobile threats include malware, Trojans, spyware, ransomware, SMS/email phishing attacks and insecure networks.

An attacker can access the following data sets that are available on a typical smartphone:

1) Corporate and personal e-mails
2). Contacts (along with their e-mail and personal addresses)
3). Banking information
4). Instant Messaging logs
5). Pictures
6). Videos
7). Credit card Information
8). Location and GPS data
9). Health information
10). Calendar and schedule information.

Thus, mobile malware is on the rise, and attackers are shifting their efforts to smartphones and tablets, putting global mobile markets at risk. This trend is continuing.

7. Evolution of SA: from Computer Viruses to Service Models

Several stages in the development of SA should be distinguished.

Among them, the main ones are the following.

Stage 1: Experiments and games (1960s - early 1980s). The first computer abuses appeared in the early 1960s and were created to teach programming to students. During this period, programmers created malicious programmes and games such as 'The Colossal Cave Adventure' and 'Eliza'. These programmes were not dangerous and did not pose a threat to computer security.

Stage 2: Viruses (mid-1980s - late 1990s). After a certain period of time, the evolution of SAs has turned them into sophisticated mechanisms used to implement attacks on information objects, the purpose of which is to destroy information or disrupt digital services of various levels of criticality. Shadow markets offer a large set of products and services for the shadow digital economy, all of which are clearly criminal in nature. In the mid-1980s, the first computer virus, Brain, was created. During this period, many viruses were created that caused computers to crash and corrupt files. In response to this, anti-virus software emerged.

Stage 3: Internet worms (1990s-early 2000s). In the late 1990s, the first Internet worm, the Morris worm, appeared. This period was also characterised by the appearance of the first Trojans - malicious programs that hide inside official software and perform undeclared malicious functions without the user's knowledge.

Stage 4: Phishing and Spyware (2000s). The early 2000s saw the emergence of new types of abuse, such as phishing - attempts to obtain a user's personal information (such as passwords and credit card numbers) through fraud, and spyware - malicious software that spies on users and can intercept their personal data.

Stage 5: Spreading malware via social media (since the 2010s). With the rise of social media, a new opportunity to spread malware has emerged. Cybercriminals began to use social media to send spam, distribute phishing messages, and infect users' computers with malware. This period also saw the emergence of a new form of computer abuse - Ransomware (for blackmail and extortion).

SAs are the basis for the following types of attacks: active, passive, internal and external. An extended list of attacks includes the following: based on the use of social engineering methods, denial of service, malware attacks, and others. The following generalised list of attacks can be proposed, including: phishing attacks, DDoS, malware infection, supply chain attacks, man-in-the-middle attacks, targeted attacks (APT), Brute-Force, identity-based attacks (logins and passwords), password spraying, internal attacks, spoofing, attacks on IoT devices, social engineering, stealth attacks, botnet, routing, etc. [32].

At present, the developments and practical experience of using SA have led to the emergence of a set of models, the main drivers of the shadow digital economy, which have contributed to the transformation of complex attack mechanisms into an accessible 'boxed' product.

It should be assumed that it is possible to further develop the family of SA on the basis of modern IT components, primarily AI, machine learning, etc., in order to automate the work on their development. It is possible to predict the emergence of AI-agents oriented on the selection of goals, adaptation of the relevant tools, etc. All this may lead, in our opinion, to the emergence of a 'self-learning' system, and this, in turn, may lead to the construction of its own full-fledged infrastructure in the form of an autonomous ecosystem.

Cybersecurity experts are increasingly interested in AI-based systems. This helps to identify new opportunities and advantages over potential adversaries. However, it should be noted that new technologies (AI, machine learning, etc.) [33], [34], [35], need to be further tested and refined [36]. The created systems should demonstrate their stability and operability, but also prove safety in new situations, not to create fundamentally new problems, how soon it will be possible to solve these problems is still unknown.

In support of this, T. O'Reilly notes: "Artificial intelligence is now at a tipping point. We may be building engines of extraordinary performance, but we are not yet building the social mechanism that will make that performance broadly applicable and broadly useful. We just hope that they evolve in some way." [37].

Individual malware with the help of AI is transformed into service models that are used in the shadow digital economy [38], [39]. Let us cite the main ones, in their composition, first of all, the following:

- Cybercrime as a Service (CaaS), an umbrella model, provides the sale/rental of criminal digital tools and services through underground marketplaces;
- Ransomware as a Service (RaaS), this is a platform where Ransomware developers provide the tools and infrastructure and 'affiliates' perform the attacks;
- DDoS as a Service (DDoSaaS), this is the leasing of capacity for DDoS attacks;
- Phishing as a Service (PhaaS), these are turnkey phishing kits (templates, landings, automation);
- Malware as a Service (MaaS), this is the rental/sale of malware (loaders, stylers, RATs, etc.) and management infrastructure;
- Money Laundering as a Service (MLaaS), these are services/networks for 'cleaning' and moving criminal proceeds (intermediaries, mules, crypto-services);
- Vulnerability Research as a Service (VRaaS), these are 'contracted' vulnerability research/technology research (legal analogue - bug bounty; in the shadows - search for further exploitation/resale);

- Data as a Service (DaaS), this is the sale of stolen data sets: PII, customer databases, 'logs' of stylers, session cookies/tokens, etc.

We consider it necessary to note the importance of further research into the shadow digital economy in this direction. It should be noted that the work of the above models is based on high automation, decentralisation and transition to new service models of interaction. There is further development of technologies and management tools related to Agentic AI (to automate operations of information theft, generation of dipfeaks, etc.), migration to encrypted messengers with preserving the balance between anonymity and convenience, transition to private blockchains to hide information about the sender, recipient and amount, use of crypto-mixers to mix cryptocurrencies of different users in order to break the connection between the original wallet and the destination address.

8. Conclusions

Malware undoubtedly has a negative technical, financial, social and emotional impact on systems and users. This is important to know because cyber security is seen as risk management and it is necessary to accurately assess how much damage a security breach can cause [12].

It is necessary to estimate the cost of malware incidents. We feel it is important to note that estimating the cost of malware is a very complex task. There are two types of costs to consider: real costs and hidden costs.

Real costs. The time that technical support staff spends troubleshooting affected computers can also be calculated. However, not all costs are so obvious.

Hidden costs. Hidden costs are costs whose impact cannot be accurately measured and may not even be known. Some businesses, such as banks and computer security companies, may suffer damage to their reputation due to a widely publicised malware incident. Regardless of the type of business, a leak of confidential information or customer data caused by malware can result in enormous damage to a company, no different than industrial espionage.

Thus, the main conclusion is the importance of malware research, the information about which is constantly being replenished in quantitative and qualitative terms. We should expect the emergence of new complex tools for counteraction.

References

1. MITRE Launches D3FEND 1.0 – A Milestone in Cybersecurity Ontology. https://www.mitre.org/news-insights/news-release/mitre-launches-d3fend-10-milestone-cybersecurity-ontology
2. Kaur J. (2019). Taxonomy of malware: virus, worms and trojan. Int. J. Res. Anal. Rev, 6(1), 192-196
3. Cohen F. Computer viruses: theory and experiments. Computers&security, 6 (1), pp. 22-35, 1987.
4. Cohen F.B. A case for benevolent viruses. all.net, 1991. [Online]. Available: https://all.net/books/tech/goodvcas.pdf.
5. Cohen F., "On the implications of computer viruses and methods of defense.," Computers & Security., vol. 7, no. 2, pp. 167-184, 1988.
6. Bezrukov N. N., KOMP'YUTERNAYA VIRUSOLOGIYA, chast 1: Obshti printsipi na funktsionirane, klasifikatsiya i katalog na naĭ-razprostranenite virusi v operatsionnata sistema MS DOS", 1990. http://read.newlibrary.ru/read/bezrukov_n_n_/page12/kompyuternaja_virusologija_ch__1.html..
7. Okhrimenko S. A., Zashtita ot kompyutŭrnite virusi., Kishinev.: Shtintsa, 1991.
8. Monnappa K. A., Learning Malware Analysis: Explore the concepts, tools, and techniques to analyze and investigate Windows malware., Packt Publishing Ltd., 2018.
9. Fites P., Johnston, P., & Kratz, M. (1988). The computer virus crisis. Van Nostrand Reinhold Co.
10. Parikka J., Digital contagions: A media archaeology of computer viruses., Peter Lang., 2007.
11. Szor P., The art of computer virus research and defense., Pearson Education., 2005.

12. Aycock J., Computer viruses and malware., Boston, MA:: Springer US., 2006.
13. Christodorescu, M., Jha, S., Maughan, D., Song, D., & Wang, C. (Eds.). (2007). Malware detection. vol. 27, Springer Science & Business Media.
14. Jajodia, S., Shakarian, P., Subrahmanian, V. S., Swarup, V., & Wang, C. (Eds.). (2015). Cyber warfare: building the scientific foundation. Springer
15. R. Burger, Computer Viruses: A high-tech disease., Abacus., 1988.
16. Jakobsson, M., & Ramzan, Z. (2008). Crimeware: understanding new attacks and defenses. Addison-Wesley Professional.
17. Sotiropoulos, J. (2024). Adversarial AI Attacks, Mitigations, and Defense Strategies: A cybersecurity professional's guide to AI attacks, threat modeling, and securing AI with MLSecOps. Packt Publishing Ltd.
18. Misra A., Dubey A. Android security: attacks and defenses. – CRC Press, 2013.
19. Hassan A. et al. ZenGuard a machine learning based zero trust framework for context aware threat mitigation using SIEM SOAR and UEBA //Scientific Reports. – 2025. Т. 15. №. 1. p. 35871.
20. Khambam S. K. R., Peta V. P., Kaluvakuri V. P. K. Augmenting SOAR with Deception Technologies for Enhanced Security and Application Response //Available at SSRN 4927248. – 2022.
21. Bontchev V. Future trends in virus writing //International Review of Law, Computers & Technology. 1997. Т. 11. №. 1. pp. 129-146.
22. Wicklund Lindroth O. Cybersecurity Ontology-The relationship between vulnerabilities, standards, legal and regulatory requirements. 2022.
23. MITRE Launches D3FEND 1.0 – A Milestone in Cybersecurity Ontology. https://www.mitre.org /news-insights/news-release/mitre-launches-d3fend-10-milestone-cybersecurity-ontology
24. Cyberdomainontology, "Cyber Domain Ontology. To develop and promote the adoption of a community developed ecosystem for standardized information representation across the cyber domain", 16.01.2025. [Online]. Available: https://cyberdomainontology.org/. [.
25. Akbar K. A. et al. The design and application of a unified ontology for cyber security //International Conference on Information Systems Security. Cham: Springer Nature Switzerland, 2023, pp. 23-41.
26. Pastuszuk J., Burek P., Księżopolski B. Cybersecurity ontology for dynamic analysis of IT systems //Procedia Computer Science. 2021. Т. 192, pp. 1011-1020.
27. Preuveneers D., Joosen W. An ontology-based cybersecurity framework for ai-enabled systems and applications //Future internet. 2024. Т. 16. №. 3. p. 69.
28. A. Guterres, UN chief outlines solutions to defeat ‘four horsemen’ threatening our global future. 01.2022. [Online]. Available: https://news.un.org/en/story/2020/01/1055791.
29. Ludwig M., Noah D. The giant black book of computer viruses. American Eagle Books, 2017.
30. G. Hinson, Cybersecurity Hyperglossary., CRC Press, 2026.
31. Burger R. Computer Viruses: A high-tech disease. – Abacus, 1988.
32. M. Smirnov, „Kompyuturni ataki - koi i kak smyata. 21 01 2026. [Online]. Available: https://bisa.ru/glavred-bdit/kompyuternyye-ataki-zachem-schitat.
33. C. Rodriguez, Generative AI foundations in Python: discover key techniques and navigate modern challenges in LLMs., Packt Publishing Ltd., 2024.
34. Sarker I. H., Furhad M. H., Nowrozy R. Ai-driven cybersecurity: an overview, security intelligence modeling and research directions //SN Computer Science. 2021. Т. 2. №. 3. p. 173.
35. Sarker I. H. AI-Driven Cybersecurity and Threat Intelligence. Springer Nature Switzerland, 2024.
36. Thakur K., Pathan A. S. K., Ismat S. Emerging ICT Technologies and Cybersecurity. Cham, Switzerland: Springer, 2023.
37. T. O’Reilly, "AI and the Next Economy," https://www.oreilly.com, 7 01 2026. [Online]. Available: https://www.oreilly.com/radar/ai-and-the-next-economy/..
38. Ohrimenco S., Borta G., Cernei V. The digital world has a long shadow //The Elgar companion to information economics. – Edward Elgar Publishing, 2024. pp. 481-504.
39. Ohrimenco S., Orlova D., Cernei V. Commercial Secret Management in Terms of Shadow Digital Economy //Business Management. 2025. Т. 35. №. 2. pp. 64-85.

Digital Transformation of Moldova's Agri-Food Sector: Risks of Creating E-Commerce Platforms for Farm Products CRM

Aurelia Turcan

Academy of Economic Studies of Moldova, Chisinau, Moldova
cce.turcan@gmail.com

Abstract. The digital transformation of Moldova's agricultural sector plays a crucial role in advancing national economic development by integrating producers into modern market dynamics, lowering transaction costs, and enhancing transparency across supply chains. Leveraging e-commerce platforms allows farmers to engage in direct sales to consumers and businesses, streamline logistics, broaden market access, and foster the growth of small and medium-sized enterprises. Digitalization also facilitates access to financial resources, encourages production upgrades, and contributes to the resilience and sustainability of rural communities. This paper explores both the benefits and the vulnerabilities associated with the digitalization of Moldova's agri-food sector. Particular attention is devoted to cyber threats — including data breaches, attacks on digital infrastructure, and forgery of electronic documents — as well as legal gaps such as insufficient regulation of online contracts, limited protection of digital signatures, and the absence of standardized approaches to product certification in virtual marketplaces. The socio-economic dimension is also addressed, focusing on risks of platform monopolization, limited digital literacy among farmers, and structural barriers to adopting online solutions. The article outlines general strategic directions for reducing these risks, emphasizing the importance of strengthening cybersecurity, refining the regulatory environment, improving farmers' digital competencies, and adopting transparent mechanisms for quality assurance. It argues that the successful digital transformation of agriculture requires an integrated approach that builds trust among market participants and ensures the long-term reliability of digital infrastructure.

Keywords. Digitalization, agricultural sector, e-commerce, cyber risks, legal regulation, online trade, Moldova.

1. Introduction

The Republic of Moldova has historically and traditionally been an agrarian country. For decades, agriculture shaped not only the economic landscape but also the nation's cultural code and the way of life for millions of people. The land has always been the main wealth, and working it - the foundation of existence.

However, today we are witnessing an alarming trend: the complete decline of the agricultural sector. Hundreds of thousands of hectares of land have been removed from cultivation, tens of thousands of farmers have gone bankrupt, and rural areas are rapidly depopulating. The share of agriculture in the country's economy is steadily declining, despite the fact that a fifth of all employed Moldovans still work the land, but with catastrophically low labor productivity.

The dynamics of the agro-sector's decline since 1990 are particularly disheartening. While in 1990 the share of agriculture in the economy was many times higher, today we observe a dramatic reduction of the sector. In the EU, over a similar period, the sector's share also decreased, but this decline was accompanied by a corresponding release of labor and an

increase in productivity. In Moldova, however, we see the preservation of archaic employment with minimal output.

Access to the internet, the basic tool of the digital economy, is available to only 74.1% of Moldovan households — 20 percentage points below the EU average of 94.1% [1, 2]. This means that a significant part of the rural population is cut off from the opportunities offered by the digital world.

Agriculture, as a strategic sector of the national economy, generates about 10% of the country's GDP. However, according to AmCham Moldova and MITP.md, the level of digital adoption in the agro-industry remains low: less than 25% of farms use digital tools for monitoring, planning, and sales, and the share of agricultural enterprises using digital sales channels does not exceed 10–12% [2]. At the same time, the IT sector accounts for 6.3% of GDP, demonstrating the potential for transferring successful digital practices to the agricultural sector.

According to the European Commission, more than 65% of Moldova's exports are directed to the European Union, where digital standards for agri-food chains are already implemented at a systemic level [3]. This makes the digitalization of Moldovan agriculture a necessary condition for increasing competitiveness in foreign markets and integrating into European digital supply chains.

In these critical conditions, digital technologies open a window of opportunity. The absence of a consistent state policy to protect local producers, combined with market openness to imports — often cheaper but not always of higher quality — has led to a systemic crisis.

A digital agri-platform operating on a direct-to-consumer (D2C) model can become the tool that allows local producers to survive and sell their products directly to consumers, bypassing the multi-level intermediaries who appropriate the lion's share of the profit. However, for this to happen, the state must create the conditions: ensure digital infrastructure, train farmers, protect platforms from cyber threats, and, most importantly, form the political will to protect the interests of its own local producers.

This paper examines existing e-commerce platforms in Moldova's agricultural sector, as well as European innovative projects that can serve as benchmarks. Special attention is paid to risk analysis — from cyber threats to market monopolization — and strategic directions are proposed for minimizing these risks within the context of national strategies and government programs.

2. Strategic framework: the Government's role in developing the agri-food sector

The foundation of all governmental initiatives in the field of agriculture and digitalization is the National Development Strategy "European Moldova 2030", approved by Law No. 315 of December 17, 2022 [4]. This document sets the development vector for the country, including the agri-food sector, in the context of European integration and the Sustainable Development Goals (SDGs).

Key provisions of the strategy relevant to the agricultural sector and digital platforms:

- Accelerated growth of labor productivity through attracting investments, technology and innovation transfer, and increasing competitiveness in the agri-food sector [5]. This creates a political mandate for investment in digital solutions, such as e-commerce platforms, which directly impact sales efficiency and farm productivity.
- Development of opportunities for innovation and entrepreneurship, improving the business environment, facilitating access to financing, and improving access to external

markets [5]. This justifies the need to create tools that help farmers enter markets, including through e-commerce support.

- Link to SDG 2.3: increasing agricultural productivity and incomes of small-scale producers through access to markets and financial services [6]. This is a direct indication of the need to support small farmers, who are the main beneficiaries of direct-to-consumer (D2C) platforms.
- Link to SDGs 8.2 and 8.3: stimulating productivity growth through technological modernization and supporting micro, small, and medium-sized enterprises [6]. This directly supports the idea of digital transformation of agribusiness and creating favorable conditions for digital startups in AgroTech.

It is important to note that the strategy does not contain direct mentions of "e-commerce platforms." However, it creates the necessary strategic framework within which such initiatives can and should develop, viewing them as a tool for achieving higher goals [7].

Specification in National Programs and Projects

Building on the "European Moldova 2030" strategy, sectoral documents are being developed, and specific governmental initiatives are being implemented. An analysis of the activities of the Ministry of Agriculture and Food Industry (MAIA) reveals a comprehensive approach to digitalizing the agricultural sector.

Table 1. Main directions of government activity [8]:

Direction	Specific Measures and Projects (2024–2025)	Significance for Agri-Platforms
Creating basic digital infrastructure	Development and implementation of information systems: Farmers' Register, RICA data collection network, Agricultural Machinery Register, AKIS system, E-ANSA	Create a transparent and verifiable database of producers, products, and supply chains. The AKIS system helps improve farmers' digital literacy
Institutional reform	Establishment of the Center for Agricultural and Consulting Advice, development of the concept for the AgroTek Park technology park	Form an ecosystem to support farmers, including digital skills training and innovation implementation
Harmonization with the EU	Participation in EU4Digital programs, preparation for IACS implementation, work on EU acquis chapters (11, 12, 13)	Orients national standards towards European ones, including traceability and quality requirements
SME and export support	ODA grant programs for digitalization and innovation (14.4 million lei for 34 companies, 21.2 million lei for 79 companies)	Direct support for companies wishing to adopt digital technologies and enter foreign markets

Food Security Strategy 2023–2030

The *Food Security Strategy 2023–2030* directly targets the structural problems of the sector. Data from the National Bureau of Statistics for 2022 shows a high level of self-sufficiency in fruits (205%) and grains (96%), but critical dependence on imports for meat

(78%), milk (63.5%), potatoes (69%), and vegetables (81%) [9]. The strategy provides for investments in milk processing, the construction of slaughterhouse infrastructure, and the creation of logistics centers for storing, sorting, and packaging vegetables and fruits [10]. E-commerce platforms can become a sales channel for local producers, helping them compete with imports and stabilize market supply.

However, declarations are not enough. Historical experience shows that without state support and protection, local producers cannot survive in an open market. This is an axiom confirmed by 30 years of post-Soviet experience. For the successful implementation of digital agri-platforms, the state must create the necessary conditions:

Table 2. Necessary conditions for the development of agri-platforms

Direction	Necessary Measures	Expected Result
Digital Infrastructure	Providing broadband internet to rural areas, supporting alternative communication technologies (LoRa)	Eliminating the digital divide, farmers' access to platforms
Digital Literacy	Creating rural innovation centers, farmer training programs, developing the AKIS system	Building human capacity for working on platforms
Legal Protection	Harmonizing legislation with EU norms, protecting producers' rights on platforms	Reducing legal risks, building trust in platforms
Cybersecurity	Developing national security standards for agri-platforms, certification	Protecting data and transactions
Financial Support	Grant programs for joining platforms, tax incentives, logistics subsidies	Lowering barriers for small producers
Institutional Support	Creating AgroTek Park, developing advisory services	Forming a support ecosystem

Source: developed by the author based on [8, 9, 10]

3. Analysis of existing agri-food platforms in Moldova and European innovative projects

The market for agri-platforms in Moldova is in an active stage of formation. Several key initiatives can be identified, differing in their business models and target audiences.

Table 3. Moldovan platforms: analysis of existing agri-food platforms

Platform / Launch Year	Type / Model / Main Focus	Key Features	Status / Dynamics
999.md 2015\ (agro-category created)	Largest national classifieds site (multi-vertical) with a special "Agriculture" section. All products, including agro (livestock, vegetables, fruits, milk, meat)	4.8 million visits/month 98% traffic from Moldova Free listings for farmers during the 2015 crisis 999Market project (partnership with the Ministry of Economy) Issues with illegal trade	Largest national platform, unique example of public-private partnership

Agromag.md 2023	B2B marketplace (multi-vendor). Agricultural inputs (seeds, fertilizers, machinery, equipment)	7,000 products 100 verified suppliers Integration with Agrobiznes.md portal	Record growth: 25K → 70K → 100K visitors (2023-2025)
Biofood.md 2022	B2C / B2B platform. Organic products	First organic platform in Moldova USAID/CNFA support MOVCA membership in IFOAM >100 products	Information on commercial success unavailable
Transoil Market 2022	B2B platform (direct purchasing). Grain crops (wheat, corn, sunflower)	Integration with TRANS OIL infrastructure (15 elevators, Giurgiulesti port) Direct sales without intermediaries Transparent real-time prices	Operational, integrated with an infrastructure giant
AgricolaHub.md 2022	Ecosystem / classifieds. Everything for agriculture + financial services	Integration with Maib (largest bank) Free listing Internal chat Information on grants, subsidies, loans	Operational, bank support ensures sustainability
Bondarenko Family 2024	Individual farm digitalization project. Own farm (cereals, orchards)	Weather station Website Remote field monitoring EU4Moldova support	Successful example, exports to the EU
Platform "Food security and food safety " 2020	Scientific-communication platform. Food security	Coordinator — Professor Rodica Sturza (ASM) Science-industry cooperation Online events (2020-2021 forum)	Operational, plays a scientific-coordination role
AgroMarket Hub 2025 (concept)	Concept (TUM project). Direct sales (D2C) for small farmers	QR traceability Shared logistics Seasonal food boxes 70% farmer willingness to participate	Only a concept, requires investment
MITP.md initiatives	Incubator / hub. Rural innovation hubs, agrotechnologies	Support for AgroTech startups Implementation of AI, automation, online marketing	Actively developing

The analysis of the platforms presented in the table reveals key features and trends in the formation of the Moldovan agri-platform market:

- **Diversity of models:** The Moldovan agri-platform market features various models — from classic B2B marketplaces (Agromag) to specialized purchasing platforms (Transoil Market), ecosystems with bank integration (AgricolaHub), and unique examples of public-private partnership (999.md with the 999Market project).
- **Dominance of 999.md:** In terms of scale and audience reach, 999.md has no equal — 4.8 million visits per month and 98% of traffic from Moldova make it the most important sales channel, including for agricultural products. The historic decision to provide free listings for farmers during the 2015 crisis demonstrates corporate social responsibility.
- **Successful niche projects:** Agromag.md shows impressive growth dynamics, indicating high demand for specialized B2B solutions in the agricultural sector.
- **Individual digitalization projects:** An example of successful digitalization of an individual farm is the family of Veronica and Veniamin Bondarenko from Slobozia village (Ștefan Vodă district) [15]. In 2024, with support from the EU4Moldova program, they received a grant for digitalization, which included installing a weather station, creating a website, and a digital platform for remote field monitoring.
- **Scientific support:** The Academy of Sciences platform plays an important coordinating role, linking science and practice, which is critical for innovation implementation.
- **Potential and gaps:** The AgroMarket Hub concept, developed at the Technical University, shows high farmer readiness for joint action (70%), but the lack of investment hinders its implementation [17].

European Innovative Projects in the Field of E-Agriculture

To understand the prospects and risks of developing e-commerce in Moldova's agri-food sector, it is useful to consider existing European projects that demonstrate advanced approaches to agricultural digitalization [18].

1) **Experimental Digital Platform LoRa in Agriculture:** The project aims to create a digital platform on an experimental farm using **narrowband radio communication (LoRa)** with internet access [18]. LoRa technology provides wireless communication with low power consumption and a long range, ideal for remote agricultural lands. For Moldova, where many rural regions face internet quality issues, LoRa could become an alternative solution for creating digital infrastructure.
2) **SOCLE Innovation AB:** This project disseminates innovative technologies among vegetable producers, providing **specific technical and economic indicators** adapted to the local production context [18]. Special emphasis is placed on resource conservation and maintaining soil organic matter. For Moldovan farmers, often distrustful of new technologies due to a lack of clear evidence of their effectiveness, this approach is critically important.
3) **Innovative Study of Robotic Mowing in Orchards:** The project tests autonomous (robotic) mowers for grass management in orchards [18]. For Moldova, with its developed horticulture (apples, plums, apricots, cherries), robotic mowing can reduce costs and decrease dependence on scarce seasonal labor.
4) **Vineyard 2.0:** A comprehensive project focused on three areas: plot monitoring and mapping (precision farming), semi-autonomous or assisted driving of machinery, and robotics in vineyards [18]. Given that viticulture and winemaking are strategic sectors

for Moldova, elements of this project could be integrated into e-commerce platforms to provide data on product origin and quality.

5) **European Research Initiatives:** The **D4AgEcol (Digitalisation for Agroecology)** project is funded by the EU to study the potential of digitalization as a factor contributing to the development of agroecological farming systems in Europe [19]. The **GreenFieldData** project, under the Marie Skłodowska-Curie Actions program, trains a new generation of researchers to address the challenges of the digital and green transition, developing IoRT-based solutions for sustainable agriculture [20].

Key Features of European Advanced Approaches

In summary, the main features of European advanced approaches to agricultural digitalization can be identified as follows:

1) **Technological diversity:** European projects cover a wide range of technologies — from simple communication systems for remote farms (LoRa) to complex robotic systems and artificial intelligence.
2) **Evidence-based approach:** Projects like SOCLE Innovation AB emphasize specific technical and economic indicators — exactly what Moldovan farmers need, as they often distrust technologies without clear evidence of effectiveness.
3) **Adaptation to local conditions:** A key principle of successful projects is adapting technologies to the local context (soils, climate, farm size), which is critically important for Moldova with its diverse conditions.
4) **Robotization of labor-intensive processes:** Projects on robotic mowing and autonomous driving show a way to solve the problem of labor shortages, which is particularly acute in Moldovan agriculture.
5) **Strategic sectors:** For Moldovan viticulture, the Vineyard 2.0 project provides a ready-made roadmap for digital transformation.
6) **Education and training:** The Hungarian experience of creating a Digital Agricultural Academy is directly applicable to solving the problem of low digital literacy among Moldovan farmers.
7) **Environmental sustainability:** Irish projects demonstrate how digital technologies can serve to preserve unique landscapes and biodiversity, which is also relevant for Moldova.

The reviewed European projects show that **the digital transformation of agriculture is not just about platforms for selling products, but also about deep technological modernization of production** [21]. For Moldova, this means that the development of e-commerce must go hand in hand with infrastructure development (LoRa), an evidence base for effectiveness (SOCLE), process automation, and enhanced traceability (Vineyard 2.0).

There is enormous potential for creating platforms for small and medium-sized producers. According to AmCham Moldova (2024), the share of agricultural enterprises using digital sales channels does not exceed 10–12% [2]. This means that the huge potential of digital sales remains unrealized.

E-commerce platforms can unite:

- small and medium-sized producers;
- farmers' cooperatives;
- returning migrants engaged in agricultural business.

Research conducted at the Technical University of Moldova shows that 70% of farmers are ready to participate in shared logistics, and almost half of consumers are willing to pay for delivery under the condition of verified product origin [17].

4. Step-by-step development of an E-Commerce platform (d2c) and risk analysis

Creating an e-commerce platform based on the D2C model for small and medium-sized agricultural producers and returning migrants requires a systematic approach. Based on an analysis of business processes and best practices, the following stages can be identified [3]:

The process of creating an e-commerce platform based on the D2C model for small and medium-sized agricultural producers includes six successive stages:

Stage 1. Situation analysis and goal setting:

- **determining the level of farm digitalization;**
- **analyzing the state of logistics infrastructure and user needs;**
- **formulating goals: increasing producer profitability, ensuring sales transparency, reducing the chain of intermediaries.**

Stage 2. Architecture development and technology selection:

- **designing a modular cloud system;**
- **integrating payment services (MAIB Pay, PayNet);**
- **developing a logistics module, analytical tools, and a user training system.**

Stage 3. Interface and content creation:

- **developing a multilingual interface (Romanian, Russian, English);**
- **adapting for mobile devices;**
- **creating and integrating training materials for farmers.**

Stage 4. Testing and training:

- **pilot launch with a limited number of users;**
- **conducting training webinars and workshops;**
- **identifying and fixing technical errors.**

Stage 5. Launch and promotion:

- **scaling the platform with support from MITP.md and AmCham.md;**
- **promotion through social media and regional agricultural associations.**

Stage 6. Monitoring and optimization:

- **analyzing performance indicators (sales volume, user satisfaction);**
- **regular feature updates;**
- **ensuring data protection and cybersecurity.**

Figure 1. **Step-by-step development of an e-commerce platform (D2C)**
(Source: developed by the author based on [3])

This approach will allow for the creation of a sustainable and secure digital ecosystem for agriculture, facilitating market expansion, increasing profitability, and integrating Moldovan farmers into international supply chains.

Successful platform implementation is impossible without a preliminary analysis of potential threats. The process of digital transformation in the agri-food sector, despite its obvious advantages, is associated with a number of risks and limitations requiring systematic analysis [22].

- **Technological risks.** The key challenge is the **low level of digital infrastructure in rural areas**. Access to broadband internet in rural Moldova remains limited, hindering the use of online platforms. Problems with software compatibility, failures in payment system integration, and insufficient reliability of cloud services can lead to data loss or platform malfunction [23].
- **Cyber and information risks.** As the volume of online transactions grows, so does the threat of **cyberattacks and leaks of confidential information** [24]. Farmer and trade platforms process data on customers, payments, and supply chains, making them targets for cybercriminals. The lack of systemic cybersecurity increases the likelihood of financial losses and loss of user trust.
- **Organizational and managerial risks.** Moldova's agricultural sector suffers from a **low level of digital literacy**, especially among small farmers and family farms. A comparative analysis shows that in Moldova, where the Agrarian University was closed in 2022, the situation with personnel training is worse than in EU countries [2]. Lack of skills in working with online platforms can lead to ineffective use of digital solutions.
- **Risks associated with seed and planting material trade.** Risks related to online seed trade deserve special attention. According to international experts, transactions between individuals (C2C) are difficult to track and are often conducted without phytosanitary certificates [25]. Material may come from small private productions without adequate disease and pest control. The likelihood of contamination of cultivated and wild plants through illegal supplies is extremely high, which could lead to economic and environmental damage.
- **Economic and market risks.** Creating and maintaining an electronic platform requires significant investment, which can be difficult for small and medium-sized producers. At the same time, high competition from foreign online markets may limit the opportunities of local participants. There is also a risk of platform monopolization by large players [26].
- **Legal and regulatory risks.** The lack of a clear regulatory framework for e-commerce in the agricultural sector also poses a threat. Moldova's e-commerce legislation is still not fully harmonized with European Union norms, creating difficulties for cross-border transactions and product certification [27].

The following measures must be implemented to neutralise the risks described:

- Development of digital infrastructure — national broadband programme.
- Improvement of digital literacy — rural innovation centres and training programmes.
- Harmonisation of legislation — acceleration of EU standards implementation.
- Strengthening cybersecurity — national standards for agricultural platforms.
- Support for small producers — grants, tax incentives, logistics subsidies.
- Ensuring traceability — QR code systems to confirm origin.
- Stimulating innovation — support for start-ups in the AgroTech sector.

5. Proposals and recommendations

Based on the analysis conducted, the following recommendations can be formulated to minimize the risks when creating e-commerce platforms in Moldova's agri-food sector.

1) **Development of Digital Infrastructure-**It is necessary to develop a national program to provide broadband internet to rural areas. Projects using LoRa technologies for remote farms, as shown in European examples, could be an alternative [18]. This also aligns with the Sustainable Development Goals (SDG 9) and the National Strategy "Moldova 2030" [4].
2) **Enhancing Digital Literacy-**The creation of rural innovation centers and training programs for farmers on working with digital platforms, CRM systems, and analytical tools is required. Programs should include not only technical skills but also the basics of cyber hygiene. The AKIS (Agricultural Knowledge and Innovation Systems) initiative, being developed by MAIA, is intended to address this task [8].
3) **Harmonization of Legislation-**It is necessary to accelerate the harmonization of national legislation with EU norms in the areas of e-commerce, data protection, and phytosanitary control. Particular attention should be paid to regulating C2C transactions with seeds and planting material, following the recommendations of the FAO Commission on Phytosanitary Measures [25]. Progress in this direction is demonstrated by MAIA's work on EU acquis chapters 11, 12, and 13 [8].
4) **Strengthening Cybersecurity -**The development and implementation of national cybersecurity standards for agri-platforms, including mandatory certification, data encryption, and personal data protection mechanisms. This must comply with the requirements of Law No. 48/2023 on cybersecurity [28].
5) **Creating Support Mechanisms for Small Producers -**Development of tax incentives and grant programs for small and medium-sized agricultural producers wishing to join digital platforms. The experience of Poland and the Baltic states shows that creating infrastructure for small exporters (consolidation centers, simplified customs procedures, special delivery tariffs) can significantly expand the export base [8].
6) **Ensuring Traceability and Transparency-**Implementing QR traceability systems to confirm product origin and quality will increase consumer trust and allow integration into European digital supply chains. The AgroMarket Hub study showed a high willingness of consumers to pay for delivery under the condition of verified product origin [17].
7) **Stimulating Innovation and Startups-**Supporting domestic startups in the agrotechnology sector, such as CityFarm, Greeno, and Solvi, which are creating new infrastructure solutions for farmers [29]. These companies are already demonstrating success internationally and could become drivers of digital transformation for the entire sector.
8) **Integration with Scientific Research -**Developing cooperation between agri-platforms and scientific institutions, such as the platforms of the Academy of Sciences of Moldova ("Food security and food safety"), to implement scientifically based approaches to digitalization [16].

6. Conclusion

The digitalization of Moldova's agri-food sector is not just a technological trend but a strategic necessity. In the long term, it can become an engine for sustainable economic growth, increasing export potential, and ensuring the country's food security.

The conducted analysis shows a comprehensive approach by the Republic of Moldova to this task. At the strategic level, the National Development Strategy "European Moldova 2030" sets goals for productivity, innovation, and support for small businesses. At the sectoral level, the Ministry of Agriculture and Food Industry is implementing specific measures to create basic digital infrastructure (Farmers' Register, RICA, E-ANSA), provide institutional support (AgroTek Park, consulting centers), and harmonize with European standards [8].

Analysis of existing platforms shows that a certain landscape of agri-platforms has already formed in Moldova — from successful B2B marketplaces (Agromag.md) to specialized niche projects (Biofood.md) and platforms integrated with large businesses (Transoil Market, AgricolaHub.md). A unique role is played by 999.md, whose historic decision to provide free listings for farmers during the 2015 crisis and partnership with the Ministry of Economy in the 999Market project demonstrate the potential of public-private cooperation in supporting local producers.

European innovative projects (LoRa, SOCLE, Vineyard 2.0) set benchmarks for further development, showing that the digital transformation of agriculture is not only about platforms for selling products but also about deep technological modernization of production [18].

However, the digitalization process is fraught with serious risks: technological (weak infrastructure), cybernetic (data leaks), organizational (low digital literacy), legal (incomplete harmonization with EU norms), and economic (risk of monopolization). It is important to note that the infrastructure being created by the state is designed to minimize these risks — ensuring the reliability of data on producers, improving digital skills, and creating favorable conditions for exports.

Successful digital transformation requires an integrated approach that fosters trust among market participants and ensures the long-term reliability of digital infrastructure. It is necessary to:

- strengthen cybersecurity;
- improve the regulatory framework;
- enhance the digital competence of farmers;
- implement transparent quality assurance mechanisms.

For Moldova, where agriculture has historically been the foundation of the economy and identity, digital transformation opens a window of opportunity. As the experience of the Bondarenko family from Slobozia shows, investment in digital technologies is not just modernization, but an expression of faith in the potential of the native land and the creation of a sustainable future for the next generations [15].

References

1. National Bureau of Statistics of the Republic of Moldova. (2024). Labour force in the Republic of Moldova: employment and unemployment in 2023. Chișinău. [and: NBS. (2025). Distribution of gross value added (GVA) by economic activity, 2024. Chișinău.]
2. AmCham Moldova, & MITP.md. (2024). Analiza sectorului digital în agricultura Republicii Moldova [Analysis of the digital sector in Moldovan agriculture]. Chișinău.
3. European Commission. (2024). Trade statistics: Republic of Moldova. [online] Available at: https://trade.ec.europa.eu/ [Accessed 28.09. 2025].
4. Law No. 315/2022 for the approval of the National Development Strategy "European Moldova 2030". (2022). Official Monitor of the Republic of Moldova, No. 409-410, art. 758.
5. Government of the Republic of Moldova. (2023). Decision No. 89 of February 28, 2023 for the approval of the National Development Plan for 2023-2025. Chișinău.
6. National Bureau of Statistics of the Republic of Moldova. (2025). SDG Report 2025_FINAL. Chișinău.

7. Cimpoieșu, A., & Șcerbacov, E. (2025). Food security: challenges and national strategies for sustainability. In: Technical-Scientific Conference of Students, Masters and PhD Students, TUM, May 14-16, 2025, Vol. III, pp. 712-716.
8. Ministry of Agriculture and Food Industry. (2024). Activity Report of the Ministry of Agriculture and Food Industry, First Half of 2024. Chișinău.
9. National Bureau of Statistics of the Republic of Moldova. (2023). Balances of food resources and their use for 2022. Chișinău.
10. Food Security Strategy of the Republic of Moldova for 2023–2030. (2022). Government of the Republic of Moldova.
11. Agrobiznes.md. (2025). Agromag.md activity report 2023-2025. [online] Available at: https://agrobiznes.md/ [Accessed 18.01. 2026].
12. MOVCA. (2022). MOVCA launched an online sales platform for organic agri-food products. [online] Available at: https://movca.md/ [Accessed 05.06. 2025].
13. TRANS OIL Group. (2022). Transoil Market Platform. [online] Available at: https://transoil.md/ [Accessed 2.02. 2025].
14. Maib. (2022). AgricolaHub.md Platform. [online] Available at: https://maib.md/ [Accessed 2.02.2025].
15. EU4Moldova. (2024). Digitalization of the Bondarenko family farm. [online] Available at: https://eu4moldova.eu/ [Accessed 4.01.2026].
16. Academy of Sciences of Moldova. (2020). Platform "Food Security and Food Safety". [online] Available at: https://asm.md/ [Accessed 2.02.2025].
17. Technical University of Moldova. (2025). Concept of the AgroMarket Hub platform. In: Student Research, Faculty of Economic Engineering and Business.
18. Raport-e-agricultura ROM. (2024). Examples of projects in the field of e-agriculture. Romania.
19. European Commission. (2024). D4AgEcol: Digitalisation for Agroecology. Horizon Europe Project.
20. European Commission. (2024). GreenFieldData: Marie Skłodowska-Curie Actions.
21. Șcerbacov, E. (2018). Aspects of food security in the Republic of Moldova. Vector European, No. 1, pp. 60-63.
22. Șcerbacov, E. (2019). Analysis of the food security through the prism of the self-insurance level of the Republic of Moldova. Vector European, No. 2, pp. 100-102.
23. ANRCETI. (2024). Report on broadband internet access in the Republic of Moldova. Chișinău.
24. STISC. (2024). Annual report on cybersecurity incidents in the Republic of Moldova. Chișinău.
25. FAO. (2023). Commission on Phytosanitary Measures: Report on online trade in seeds and planting material. Rome.
26. Oberșt, A. (2021). Development of food safety culture – the imperative of performant leadership. In: Competitiveness and sustainable development, 3rd Edition, pp. 153-159.
27. Law No. 48/2023 on cybersecurity of the Republic of Moldova. (2023). Official Monitor of the Republic of Moldova.
28. MITP.md. (2024). Innovative startups in the agro-technology sector of the Republic of Moldova. Chișinău.

E-Commerce: Innovative Tools for Digitalizing Trade in Moldova

Aurelia Turcan, Lucia Gujuman

Academy of Economic Studies of Moldova, Chisinau, Moldova
cce.turcan@gmail.com, gujuman.lucia@ase.md

Abstract. Global digitalization has transformed online commerce into a key factor for economic development. For the Republic of Moldova, this shift is a powerful opportunity to open up new markets, boost entrepreneurship, and connect local producers to global supply chains. This article takes a close look at the innovative tools driving modern e-commerce, from online marketplaces and digital payment systems to automated logistics and smart marketing tech. We explore how these tools are making transactions more transparent, cutting costs, and speeding up buying and selling. A special focus is placed on the impact for small and medium-sized businesses. The analysis also addresses the hurdles that slow down digital adoption in Moldovan trade, such as gaps in infrastructure and digital skills. The piece argues that for Moldova to truly succeed in digitizing its trade sector, a well-rounded strategy is essential. This includes modernizing regulations, strengthening cybersecurity, and creating strong support programs for entrepreneurs. In conclusion, there is a clear need for more research and the active adoption of innovative practices to fast-track the Moldovan economy's integration into the global digital marketplace.

Keywords: e-commerce, innovation, digitalization of trade, Moldova, online platforms, digital payments, logistics.

1. Introduction

Global digitalization has transformed online commerce into a key factor for economic development. For the Republic of Moldova, as for many countries with transition economies, this transformation opens up powerful opportunities: access to new markets, stimulation of entrepreneurship, and integration of local producers into global value chains. At the same time, the process of digital adaptation faces serious challenges, including infrastructure constraints, a digital skills deficit, and the need to harmonize legislation with the norms of the European Union (EU). The course towards European integration, adopted by Moldova, gives this issue particular urgency, requiring not only technological but also regulatory modernization of the e-commerce sector. However, digitalization brings not only benefits but also significant environmental costs, making the search for pathways to a sustainable digital economy an imperative.

From the first experiments with Electronic Data Interchange (EDI) in the 1960s to giants like Amazon and Alibaba, e-commerce has come a long way. Today, it is not just a sales channel but a complex ecosystem operating in a VUCA world (Volatility, Uncertainty, Complexity, Ambiguity). This reality, characterized by volatility, uncertainty, complexity, and ambiguity, exerts colossal pressure on business. The response to this challenge is digital transformation, whose goal is not just technology implementation but value creation. The main value and core digital reality becomes Customer Experience. Competition revolves around

creating a seamless, personalized, and contextual customer journey from the first touchpoint to after-sales service.

There is a gap between the rapid development of business practices in e-commerce and the academic understanding of these processes. Small and medium-sized enterprises (SMEs) face the greatest difficulties trying to adapt to the dual transition — simultaneous digitalization and the shift towards sustainable development. The purpose of this article is to synthesize existing knowledge and present a comprehensive picture of the role of innovation in e-commerce. This includes a classification of innovation types, an analysis of the impact of key disruptive technologies, their connection to sustainable development, and ultimately, a proposal for a practical framework for implementing innovations in SMEs, based on the foundational work of Purcarea (2024).

Objective of the Study. The aim of this article is a comprehensive analysis of the role of innovation in the transformation of electronic commerce in Moldova, examining their evolution from a technological driver to a tool for sustainable development. The work sets the following tasks: (1) to present a review and analysis of scientific publications on the topic of innovation in e-commerce; (2) to define the main concepts and develop a typology of digital trade; (3) to structure and analyze the key innovations applied in the Moldovan e-commerce sector; (4) to assess the alignment of this development with the UN Sustainable Development Goals (SDGs) in the context of European integration; (5) to propose a practical system for evaluating innovations to choose the optimal strategy.

2. Theoretical Foundations: Innovation as the Engine of the Fourth Industrial Revolution

From Industry 4.0 to Industry 5.0 and the Business2Robot2Consumer (B2R2C) Concept

The concept of the Fourth Industrial Revolution (4IR), initiated in Germany as "Industry 4.0," describes the shift caused by the fusion of digital, physical, and biological worlds. It signifies the digital transformation of industry based on cyber-physical systems, the Internet of Things (IoT), and cloud computing. However, today we are already talking about Industry 5.0, which expands this focus. It prioritizes not only efficiency but also sustainability, human-centricity, and resilience (the ability to withstand and quickly recover from shocks). Industry 5.0 is intended to lead both the digital and "green" transitions, ensuring prosperity within planetary boundaries.

In this context, robotization is having an increasing impact on e-commerce. From warehouse automation to drones for the "last mile," robots are becoming an integral part of the value chain. This gives rise to a new concept — **Business2Robot2Consumer (B2R2C)**. Marketers and strategists need to consider that their target audience is increasingly not only people but also algorithms, smart speakers, and robot assistants that make purchases on behalf of a person or on their behalf.

Digital Disruption and Business Model Transformation

Digital disruption occurs when new players with innovative business models challenge established "incumbents." According to Clayton Christensen's theory of disruptive innovation, new competitors often enter the market with simpler and cheaper solutions, targeting overlooked or underserved segments. They use a disruptive innovation strategy, while incumbents focus on sustaining innovations, improving products for their most demanding customers, and often miss the threat from below.

Key to understanding digital business is the "iron triangle," describing three interconnected perspectives of any digital model:

- **Who? (participants):** creating and exchanging value in the network.
- **What forms? (value):** how value is specifically created and delivered.
- **Sustainability? (durability):** how the model maintains competitiveness over time.

Achieving a balance between these three elements is the main strategic task for any company in the era of digital disruption

3. Classification and Role of Innovations in E-commerce

Understanding the multifaceted nature of innovation is key to analyzing the transformation of e-commerce. Based on the comprehensive study by Faccia et al. (2023), AmCham (2024), Vasilache A. (2004). which provides a foundational typology for understanding technological change, we can identify and analyze various types of innovations, each playing a unique role in the digitalization of trade, possessing its own advantages, and carrying certain risks.

1). Disruptive Innovations

- **Role and Impact:** Disruptive innovations do not just improve the existing — they create new markets and value networks, often starting with simple and affordable solutions in niches ignored by established players. In e-commerce, their role is fundamental: they democratize market access for both sellers and buyers.
- **Pros:** Creation of entirely new business models (e.g., marketplaces uniting millions of sellers and buyers). Stimulation of competition and price reduction for the end consumer. Expansion of consumer choice and reach to a global level.
- **Cons / Risks:** High degree of uncertainty and risk of failure. Destabilization of established markets, potentially leading to the bankruptcy of traditional players and job loss. Require significant changes in legislation and regulation.
- **Examples:** The emergence of giants like Amazon and Alibaba completely changed the retail landscape. A more local example for Moldova is the launch of the 999.md/ro/marketplace, which, although still in its initial stage of development, is already creating a new ecosystem for local sellers and buyers, competing with traditional sales channels.

2). Sustaining Innovations

- **Role and Impact:** This is the main driver for the continuous improvement of existing products and services. They aim to meet the needs of current customers, making products better, faster, more convenient, and more reliable. In digital trade, this is the engine for improving the quality of customer experience.
- **Pros:** Predictability and relatively low risk. Allow companies to retain and strengthen their market share. Increase customer loyalty through continuous improvement of familiar services.
- **Cons / Risks:** A focus on sustaining innovations can lead a company to miss the emergence of disruptive technologies and new market niches (the "incumbent's trap"). They do not create new markets, only develop existing ones.
- **Examples:** Implementing faster delivery (e.g., expanding the geography and speed of courier services like Nova Poshta in Moldova), improving the interface of an online store for easier navigation, optimizing the mobile version of the site, implementing loyalty programs.

3). Breakthrough Innovations

- **Role and Impact:** These are innovations that create new technological trajectories and often lie at the intersection of different fields. They do not just improve but

fundamentally change the way a need is met, making possible what was previously impossible

- **Pros:** Ability to create entirely new product and service categories. Give the companies that first implement them a huge competitive advantage. Can dramatically increase efficiency and reduce costs.
- **Cons / Risks:** High development cost and complexity. Require significant changes in behavior and skills from companies and consumers. Often face infrastructure unreadiness or regulatory barriers.
- **Examples:** Creation of digital wallets and instant payment systems (like the new MIA system in Moldova), platform solutions uniting vast ecosystems (Amazon Web Services as a platform for thousands of other businesses)

4). Frontier Innovations

- **Role and Impact:** These are innovations at the cutting edge of scientific and technological progress. They are at the stage of experiments and research, and their potential is yet to be revealed. For digital trade, they define the contours of the future.
- **Pros:** Potential for radical efficiency gains (e.g., autonomous drone delivery), personalization (AI), transparency (blockchain). Open access to new, previously non-existent data and analysis opportunities.
- **Cons / Risks:** Extremely high degree of uncertainty and investment risk. Technologies may prove commercially unviable or face insurmountable ethical or regulatory obstacles. Require unique competencies.
- **Examples:** Using the Internet of Things (IoT) to create fully automated supply chains, applying 5G for augmented reality (AR) technologies in shopping, experimenting with quantum computing to optimize logistics.

5). Radical Innovations

- **Role and Impact:** Unlike breakthrough innovations, radical innovations create entirely new products for the *existing* market. They replace old technologies, completely supplanting them.
- **Pros:** Allow a company to achieve a dominant market position if it can quickly capture share by displacing old products. Stimulate technological progress in the industry.
- **Cons / Risks:** High risk of market rejection if the new product is too complex or does not solve real consumer problems. Require significant investment in marketing and consumer education.
- **Examples:** The shift from traditional retail to online marketplaces, the emergence of VR stores as a new form of brand interaction, the rise of social commerce through Instagram and Facebook, which created a new sales channel for millions of small entrepreneurs.

6). Architectural Innovations

- **Role and Impact:** They restructure the internal architecture of a system, leaving its components unchanged. This is a change in how the parts of the system interact with each other. In digital trade, this is the foundation for flexibility and scalability.
- **Pros:** Allow quick adaptation to changing market demands. Simplify the integration of new technologies and services. Increase system stability and reliability.
- **Cons / Risks:** Complexity and high cost of restructuring established architectures. Require a highly qualified development team. May temporarily reduce system performance during the transition period.

- **Examples:** Transitioning from a monolithic platform to a microservices architecture, allowing individual parts of an online store to be updated and scaled independently. Implementing a headless CMS, which separates content management from its presentation, allowing the same content to be used on the website, mobile app, and social networks.

7). Modular Innovations

- **Role and Impact:** They affect individual system components, replacing them with new, more advanced ones, but leaving the overall architecture unchanged. This is a way of gradual but deep modernization.
- **Pros:** Lower risk and cost compared to architectural changes. Allow quick implementation of new functions by replacing an "old module with a new one." Simplify system maintenance and updates.
- **Cons / Risks:** Potential compatibility issues of the new module with the old ones. Risk of vendor lock-in if the module is too tightly integrated with a single supplier.
- **Examples:** Replacing an old payment gateway with a new, more functional and secure one via API. Implementing a new AI-based product recommendation system without changing the rest of the store's functionality.

8). Incremental Innovations

- **Role and Impact:** This is a continuous stream of small, gradual improvements. They form the basis of daily work on optimizing products and processes. Their cumulative effect over time can be enormous
- **Pros:** Low risk and cost. Easy to test and implement. Allow continuous learning from user feedback and data. Create a culture of continuous improvement.
- **Cons / Risks:** Individually, each innovation has a small effect. Can create an illusion of development when a company misses the need for more fundamental, radical changes.
- **Examples:** Regular A/B testing of headlines and buttons on a website to increase conversion. Continuous optimization of product description texts based on search queries. Minor improvements to the checkout interface to reduce shopping cart abandonment.

9). Open vs. Closed Innovations

- **Role and Impact:** These two paradigms define the *source* of innovation. Open innovation involves the active use of external ideas, partnerships, and technologies. Closed innovation relies on internal research and development to protect intellectual property.
- **Pros (Open):** Access to a wider pool of ideas and talent. Accelerated development by using ready-made solutions. Reduced R&D costs. Creation of ecosystems and network effects.
- **Cons (Open):** Risk of losing control over key technologies. Complexity of managing intellectual property rights. Dependence on external partners.
- **Pros (Closed):** Full control and protection of unique developments. Possibility of creating a sustainable long-term competitive advantage.
- **Cons (Closed):** High costs and risks of internal R&D. Limitations of own expertise and perspective. Risk of "reinventing the wheel" and lagging behind the market.
- **Examples (Open):** Using Shopify to quickly create an online store, integrating payment services like PayPal or local PSPs, participating in partner programs and hackathons.

- Examples (Closed): Developing a proprietary e-commerce platform with unique functionality, creating a unique recommendation algorithm that the company does not disclose to competitors.

The detailed classification of innovations carried out is not just an academic exercise. It is a powerful analytical tool that provides businesses, regulators, and researchers with several important advantages.

What does this knowledge give us?

- **Strategic Planning and Positioning:** Understanding the different types of innovations allows a company to clearly define its strategy. Will it be a "pioneer" implementing frontier technologies (high risk, high potential reward), or a "fast follower" improving existing solutions (incremental/sustaining innovations) with lower risk? This knowledge helps answer the question: "Who are we in this market and where do we want to go?"
- **Resource Optimization and Risk Management:** Each type of innovation requires different resources and carries different risks. Understanding this allows for more efficient allocation of R&D budgets and investments. A company can balance its innovation portfolio, investing simultaneously in "risky but promising" breakthrough projects and in "reliable but evolutionary" sustaining improvements.
- **Competitive Analysis:** Using this typology as a "map," one can analyze the actions of competitors. What type of innovation is your main rival using? Are they building an ecosystem (disruptive) or just improving their website (sustaining)? This allows anticipating their next steps and developing an effective response.
- **Assessing Market and Ecosystem Maturity:** At the country or region level (as in the case of Moldova), this classification helps assess the maturity of the digital ecosystem. The predominance of incremental innovations may indicate stagnation and a lack of breakthrough ideas. Conversely, the emergence of even isolated disruptive or architectural projects (like the launch of the 999.md marketplace or the introduction of the national MIA instant payment system) is a strong signal of the market transitioning to a new level (AmCham, 2024).

How to evaluate which type of innovation is "better"?

There is no universal answer to this question. The "best" innovation is the one that best aligns with the company's current goals, resources, and context. For an objective assessment, we propose using a multi-criteria approach, considering innovations not as isolated phenomena, but as elements of a single portfolio.

How to use this table?

- **Define your goal.** What is more important for you: quickly capturing market share with minimal risk (focus on incremental/sustaining innovations) or creating a new market with potential for super-profits, but also high risk (focus on disruptive/breakthrough)?
- **Assess your resources.** Do you have the capital for long-term research (frontier innovations) and is your team ready for high uncertainty? Or is the budget limited and quick returns needed (incremental)?
- **Analyze the context.** Is your market mature? How intense is the competition? What are your customers' expectations?
- **Aim for balance.** The most sustainable strategy is not betting on one type of innovation, but creating a balanced portfolio. For example: 70% of resources on sustaining and incremental innovations (provide current income and loyalty), 20% on architectural and breakthrough (prepare the company for tomorrow), and 10% on disruptive and frontier experiments (seek opportunities for the day after tomorrow).

Table 1. Comparative analysis of innovation types for choosing the optimal strategy

Evaluation Criterion / Innovation Type	**Disruptive**	**Sustaining**	**Breakthrough**	**Frontier**	**Architectural**	**Incremental**
Market Growth Potential	Extremely High (creating new markets)	Low / Medium (improving share in existing market)	Very High (capturing new niches)	Uncertain, potentially revolutionary	High (improving efficiency and flexibility)	Low
Risk Level	Very High	Low	High	Critically High	Medium	Very Low
Required Investment	High	Low / Medium	Very High	Extremely High	High	Low
Time to Result	Long-term	Short-term	Medium / Long-term	Long-term (experiments)	Medium-term	Short-term
Impact on Customer Experience (CX)	Revolutionary (new experience)	Evolutionary (improving current experience)	Radical change in experience	Creation of fundamentally new experience	Indirect (via speed and reliability)	Small, but continuous improvement
Impact on Operational Efficiency	High (new business models)	Medium (process optimization)	High	Potentially very high	Very High (system flexibility)	Low
Applicability for Moldova (example)	999.md/ marketplace (creating local C2C/B2C market)	Improving interface and delivery speed at emag.ro, darwin.md	MIA instant payment system	Pilot projects for delivery drones (in perspective)	Banks transitioning to microservices architecture for their apps	A/B testing of landing pages

Knowing the typology of innovations and being able to evaluate them against key criteria transforms the innovation process from a chaotic search for "something new" into a manageable strategic discipline. For Moldova, on the threshold of active European integration, this approach allows not only copying best global practices but also creating its own unique competitive advantages, optimally using limited resources and adapting to the specifics of the local market.

4. Innovations in Moldova's E-commerce: Structure and Dynamics

The Moldovan e-commerce market, according to the detailed report by AmCham Moldova (2024), shows steady, though uneven, growth.

Table 2. Key indicators of the Moldovan e-commerce market (2023)

Indicator	**Value (2023)**	**Notes**
Total market volume (goods + services)	EUR 358.6 million	Growth of 15.2% compared to 2022
Goods market (B2C goods)	EUR 229.1 million	63.9% of the market, growth of 18.8%
Services market (B2C services)	EUR 129.5 million	36.1% of the market, growth of 9.4%
E-commerce share in retail trade	5.60%	Indicates significant growth potential
Number of active e-shops (goods)	470-520	Estimate based on analysis of 420 stores
Number of active online platforms (services)	~480	Platforms accepting online payments
Active online audience (18-65 years, urban)	~327,600 people	~20% of the adult population
Average monthly e-commerce spending	EUR 91.4	Including EUR 58.4 on goods

[Source: AmCham, 2024]

The analysis of the Moldovan e-commerce market conducted by AmCham Moldova (2024) identified four key groups of innovations that have become growth drivers for the sector. By the beginning of 2026, these directions not only maintained their relevance but also received significant development.

1. Development of e-hypermarkets and the ecosystem approach

Projects like Smart, PandaShop, and Smadshop laid the foundation for the formation of a mature online retail market. By early 2026, these platforms evolved from simple online stores to full-fledged ecosystems. For example, **Pandashop.md** is actively developing its mobile application, offering not only an expanded assortment (from electronics to home goods, auto accessories, and office supplies) but also integrated solutions: convenient payment methods (including online, cash on delivery, bank transfer), fast delivery throughout Moldova, and a personalized approach to customers.

The **999.md marketplace**, in turn, continues to strengthen its position, offering goods unique to the Moldovan market, including international brands (e.g., IKEA furniture through local partners), which creates a competitive advantage over traditional offline retail and expands consumer choice .

2. Food delivery services (q-commerce) and hyper-personalization

The Straus and Glovo projects, launched during the pandemic, had by 2026 become an integral part of the daily lives of Moldovan consumers, radically changing consumption habits in the ready-meal segment. The **Straus.md** service, by early 2026, had significantly expanded its presence, uniting around 180 establishments with diverse cuisines — from classic burgers and pizza to Mexican, Thai, Indian, and traditional Moldovan cuisine. Key innovations included:

- Personalized recommendations based on order history.
- Process transparency: the ability to track order status in real-time, rate restaurants, and leave comments on the composition of dishes.
- Technical excellence: regular application updates aimed at increasing loading speed and operational stability, as well as bug fixes.

3. Development of "last mile" logistics

The expansion of operators like **Fan Courier** and **Nova Poshta** continues to play a critical role in the development of e-commerce. By 2026, their coverage networks became even denser, covering not only large cities but also remote settlements. Competition among logistics operators led to:

- Reduced delivery costs for end consumers and retailers.
- Shortened delivery times (emergence of express delivery services on the day of order).
- Development of a network of parcel lockers and pickup points, increasing convenience for buyers who can collect goods at any convenient time.

5. Security Evolution: From 3-D Secure Implementation to Dynamic Authentication and European Standards

If in the early 2000s, the implementation of the **3-D Secure** protocol (Verified by Visa, Mastercard SecureCode) by Moldovan banks and the Romcard processing center was a fundamental innovation that laid the foundation for secure online payments (Vasilache, 2004), then by 2026 this system has evolved to a qualitatively new level.

Key changes by 2026 include:

- **Transition to dynamic authentication (3D Secure 2.0).** The static password, vulnerable to phishing, has given way to dynamic one-time passwords (OTP). Moldova Agroindbank (MAIB) was the first in Moldova to implement this standard, making the use of stolen card data online practically impossible (Turcan A., 2024b).
- **Expansion of authentication methods.** Modern 3DS implies multi-factor authentication: confirming purchases via biometrics (Touch ID / Face ID) in banking mobile apps, push notifications, and frictionless authentication, where the system analyzes over 100 transaction parameters and confirms the payment in the background without bothering the user (Vasilache, 2004).
- **Harmonization with European regulatory standards.** By 2026, the Moldovan financial sector is integrating into the European regulatory field:
  - The implementation of Strong Customer Authentication (SCA) requirements (analogous to the European PSD2 directive) makes strong authentication mandatory for the vast majority of online payments. Transactions without 3DS verification may be declined by banks for security reasons (Turcan A., 2024a).
  - The National Bank of Moldova (NBM) is actively implementing the European Digital Operational Resilience Act (DORA), which establishes uniform

requirements for ICT risk management and digital resilience testing for all financial organizations (Turcan A., 2024a).

- **Mandatory standard for successful transactions.** By 2026, 3DS support has become not just a "feature," but a necessary condition for a high level of payment conversion. Many banks may "soft decline" transactions from cards not ready for 3DS, stimulating all market participants to adopt modern standards (Vasilache, 2004).

Thus, if the initial implementation of 3-D Secure was the foundation of trust, then by 2026 we are witnessing the formation of a **holistic, multi-level security ecosystem**, based on dynamic authentication, biometrics, strict regulatory requirements (SCA, DORA), and harmonization with European standards.

At the same time, the Moldovan market retains a number of specific features. The main one is the dominance of cash or card on delivery, which is associated with a relatively low level of consumer trust. This practice entails additional operating costs for e-retailers (AmCham, 2024).

- **Innovations as a Catalyst for Sustainable Development**

The Dual Transition: Synergy of Digitalization and Sustainability

Sustainable development is the path, and sustainability is the ultimate goal, requiring a balance of economic, social, and environmental factors. Digitalization, in turn, offers many tools to achieve this goal.

The dual transition means that companies must simultaneously manage digital transformation and the integration of ESG (Environmental, Social, Governance) principles into their operations. Supply chain responsibility becomes critically important: consumers and regulators demand transparency in the origin of goods, ethical production, and minimization of the environmental footprint — a challenge highlighted in global discussions on digital sustainability.

The development of e-commerce directly correlates with the achievement of several UN Sustainable Development Goals, as outlined in key international reports (UNCTAD, 2024):

- **SDG 8 – Decent Work and Economic Growth:** Creation of new jobs in the digital sector, support for SMEs through access to online markets — *Package of Measures for De-bureaucratization, Approved in December 2025 [cited in Moldpres, 2026];.
- **SDG 9 – Industry, Innovation, and Infrastructure:** Development of digital and logistics infrastructure, introduction of innovations in trade processes — *Government Decision on the Digital Transformation Program for the Energy Sector for 2026–2030 [Approved in October 2025]*;
- **SDG 4 – Quality Education:** Digital transformation of education is critically important for training qualified personnel capable of working in the new digital reality and safely using its capabilities — *Digital Transformation Strategy of the Republic of Moldova for 2023–2030 (Draft)*.
- **SDG 10 – Reduced Inequalities:** Digital transformation strategies should aim to reduce the digital divide between urban and rural residents, ensuring equal access to services and opportunities — *National Development Strategy "European Moldova" 2030*;

The Environmental Footprint of Breakthrough Technologies and the Path to Sustainability

The relationship between digitalization and environmental sustainability is complex and multifaceted. Global research emphasizes that digitalization has a significant environmental footprint throughout the entire technology life cycle: from raw material extraction to the disposal of electronic waste. This makes the search for pathways to a sustainable digital economy an imperative (UNCTAD, 2024).

Impact of key breakthrough technologies on sustainable development:

- **Artificial Intelligence (AI) and Machine Learning (ML):** They permeate all aspects of e-commerce. AI enables personalized recommendations, optimizes real-time pricing, forecasts demand for inventory management, and automates customer support with chatbots. However, AI requires significant computing power, leading to increased energy and water consumption in data centers — a growing concern addressed in international sustainability frameworks.
- **Immersive Reality Technologies (AR/VR/XR):** They blur the line between online and offline. AR fitting rooms allow you to "try on" clothes or see how furniture would look in a room. Virtual showrooms create a new level of engagement, transforming e-commerce into iCommerce (immersive commerce).
- **Web3 Technologies (Blockchain, NFT, smart contracts):** They offer a decentralized approach, increasing transparency and trust. Blockchain can be used to track the authenticity of goods (luxury, art), and smart contracts for automating payments and managing supply chains. New NFT-based loyalty models can create deeper connections with customers. However, cryptocurrency mining is an extremely energy-intensive process, a fact well-documented in analyses of digitalization's environmental impact.
- **Generative AI (GenAI):** Became a real breakthrough in 2023. Tools like ChatGPT and DALL-E are revolutionizing content creation. They allow instant generation of thousands of unique product descriptions, creation of visuals for marketing campaigns, development of new product designs, and provision of personalized real-time customer support (Faccia et al., 2023).

Innovations in the following areas are necessary for the transition to sustainable development:

- **Logistics and the "last mile":** Optimizing delivery routes, using electric vehicles and bicycles, pickup points (click&collect) instead of courier delivery for every single order (AmCham, 2024).
- **Packaging:** Switching to recyclable and biodegradable materials, optimizing packaging size, eliminating excess plastic (AmCham, 2024).
- **Circular economy (economie circulară):** Marketplaces can become drivers of the circular economy, actively promoting the sale of used, refurbished goods and creating platforms for buybacks — a key recommendation for aligning digital growth with sustainability goals (UNCTAD, 2024).
- **Integrating sustainability into the business model:** Successful companies no longer view sustainability as a separate PR campaign. They integrate it into the core of their strategy. For analyzing readiness to adopt such innovations, theoretical frameworks like TAM (perceived usefulness and ease of use), TPB (influence of social norms and perceived control), and TOE (technological, organizational, and environmental context) are used.

6. Practical Guide: From Theory to Action for Small and Medium Enterprises (SMEs)

Based on the synthesis of theoretical knowledge and empirical data, an authorial tool was developed by Purcarea (2024) — the **"Decision Pyramid for Using Breakthrough Technologies by SMEs in E-commerce"**. It consists of six sequential levels (P1-P6).

- **P1: Combining Digital Transformation with Sustainable Development.** Foundational level. The SME must clearly define how it combines digital initiatives with sustainable development goals, assess its environmental footprint, and choose the right approach to transformation (for a new or already operating business). This also requires alignment with national strategies, such as the Draft Digital Transformation Strategy of the Republic of Moldova for 2023-2030.
- **P2: Fundamental Choice of E-commerce Architecture.** A critical decision affecting scalability and flexibility. This involves choosing between monolithic architecture, microservices, or headless solutions, and correctly selecting tools for back-end, front-end, checkout, and marketplace functionality, based on principles outlined by Vasilache (2004).
- **P3: Developing the Capacity for Technological Transformation.** This involves assessing the current IT infrastructure, defining user profiles, implementing flexible methodologies (Agile/DevOps), and ensuring cybersecurity. Understanding how specific technologies (AI, IoT, Blockchain) can be applied to increase productivity is essential for modern e-commerce businesses — a key insight from global digital economy analyses.
- **P4: Understanding Customer Motivations for Adopting New Technologies.** An in-depth study of the target audience is necessary: what is more important to them — convenience, personalization, innovativeness, price, or social responsibility? This also includes developing educational materials for customers (tutorials, knowledge bases), based on insights from the Moldovan market (AmCham, 2024).
- **P5: Using Generative AI.** This level focuses on the most current trend. Implementing GenAI for automating content creation (descriptions, design), improving customer service (smart chatbots), personalization, and assortment management across all channels (e-commerce, m-commerce, social, voice) represents a significant opportunity for SMEs to enhance their competitiveness (Faccia et al., 2023).
- **P6: Customer Centricity and Continuous Experience Improvement.** The pinnacle of the pyramid. This involves using all previous levels to create a seamless, personalized, and immersive customer experience. Continuous collection of feedback and data analysis for improvement directly leads to increased loyalty and profitability, a key finding in both theoretical frameworks and the Moldovan market analysis (AmCham, 2024).

7. Conclusion

The conducted analysis allows drawing several key conclusions. Innovations in modern e-commerce are not just about implementing individual technologies, but a comprehensive transformation of the business model, corporate culture, and operational processes. Success in the new VUCA reality requires companies, especially SMEs, to simultaneously manage digital transformation and the shift towards sustainable development. Customer experience, enriched by immersive technologies and AI-based personalization, becomes the central element of competitiveness.

For Moldova, the path to sustainable development of digital trade is inextricably linked with European integration. The legal and strategic framework being created by the country in close cooperation with the EU provides a solid foundation for this development. Harmonization with European norms not only facilitates cross-border operations for businesses and citizens but also ensures a high level of consumer protection, stimulates investment and innovation. However, the success of this path depends on the ability of the state and businesses to jointly manage the risks associated with cybersecurity and rapid technological changes, as well as on investments in human capital and digital education (Turcan A., 2024a, 2024b).

The proposed "Decision Pyramid" developed by Purcarea (2024) represents a practical tool that will help small and medium-sized enterprises structure this complex path and effectively use breakthrough technologies to enhance their competitiveness.

This work opens up several perspectives for further research. Firstly, an in-depth study of the long-term impact of generative AI on all aspects of e-commerce — from value creation to ethical dilemmas - is needed. Secondly, the development of universal and accurate methodologies for measuring the environmental and social footprint of digital commerce is an urgent issue, which would allow assessing real progress towards sustainable development — a challenge recognized in international policy discussions. Thirdly, the adaptation strategy of small and medium-sized businesses to the rapidly changing technological landscape, especially under resource constraints, requires separate analysis. Finally, the future lies in studying the role of emerging platforms such as the metaverse and super-apps, and their potential in shaping the next generation of digital customer experience.

References

1. AmCham Moldova. (2024). Piata comerțului electronic în Republica Moldova în 2024. AmCham Moldova.
2. Faccia, A., Le Roux, C. L., & Pandey, V. (2023). Innovation and E-Commerce Models, the Technology Catalysts for Sustainable Development: The Emirate of Dubai Case Study. Sustainability, 15(4), 3419.
3. Iancu, A. (Coord.). (2004). Dezvoltarea economică a României. Editura Academiei Române.
4. Ioan-Franc, V. (2000). *Comisia de Fundamentare a Strategiei Naţionale de Dezvoltare Economică. Din lucrările Comisiei de la Snagov – martie-iunie 1995*. Academia Română, Centrul de Informare şi Documentare Economică.
5. Ogden, C. K., & Richards, I. A. (1923). The Meaning of Meaning: A Study of the Influence of Language upon Thought and of the Science of Symbolism. Harcourt, Brace & World.
6. Purcarea, I.-M. (2024). Impactul tehnologiilor disruptive asupra e-commerce. Academia Română, Institutul Național de Cercetări Economice „Costin C. Kirițescu".
7. Turcan, A. (2024a). Правовое регулирование электронной коммерции в Республике Молдова (Legal regulation of E-Commerce in the Republic of Moldova). New Knowledge. Journal of science, XIII(4).
8. Turcan, A. (2024b). Digitalization and innovation in Moldova's E-commerce payment systems: opportunities and risks. New Knowledge. Journal of science, XIII(4).
9. United Nations Conference on Trade and Development (UNCTAD). (2024). Доклад о цифровой экономике 2024: Формирование экологически устойчивого и инклюзивного цифрового будущего (Обзор). UNCTAD/DER/2024 (Overview).
10. Vasilache, D. (2004). Comerţul electronic, eComerţ. In Plăți Electronice. (Textbook).

Using Artificial Intelligence to Ensure Information Security of Organizations: Possibilities, Problems and Prospects

Olga V. Pugacheva

Francisk Skorina Gomel State University, Gomel, Belarus
OPugacheva@gsu.by

Abstract. The article substantiates the relevance of studying the possibilities of using artificial intelligence in the information security of organizations, which is determined by the need to increase the level of protection, improve response to incidents and reduce the risks associated with cyberattacks. For this purpose, statistics on global threats to information security are analyzed based on the number of incidents in 2023 and 2024. On this basis, the main tasks of AI technologies in ensuring the information security of organizations and their content are formulated. The article analyzes the capabilities of AI solutions at individual stages of information security: warning and forecasting threats; penetration testing and attack simulation; detection of malicious and abnormal activity; processing of security events; recognition of phishing, unwanted content, dangerous sites, etc. As a result of the analysis, several levels of AI technology use in ensuring information security are identified: AI is already being used, AI can be used, the use of AI is not justified or will not bring significant benefits. The study highlights such prospects for using AI in the field of information security as the creation of autonomous systems based on AI and modeling a cyber range, and also considers problems associated with the complexity of implementing AI technologies in this area.

Keywords. Information security, threats, artificial intelligence, technologies, organizations

1. Introduction

The relevance of studying the possibilities of using artificial intelligence (AI) in the information security of organizations is determined by the need to increase the level of protection, improve response to incidents and reduce risks associated with cyber threats.

The possibilities of artificial intelligence technology in ensuring information security of organizations are determined by many factors, the main ones of which are [1, 2, 3]:

- increase in cyber threats. With the rise of digitalization and the transition to remote work, the number of cyberattacks has increased significantly. Hackers are using increasingly sophisticated methods, which requires organizations to take new approaches to protecting information.

- Big data processing. AI is able to process and analyze huge amounts of data in real time, which allows identifying anomalies and potential threats faster and more effectively than traditional methods.

- process automation. Using AI in information security allows you to automate routine tasks such as system monitoring, log analysis, and incident response, freeing up resources for more strategic tasks.

- predictive analytics. Machine learning algorithms can predict potential threats based on historical data, allowing organizations to defend against attacks.

- improving threat detection. AI can improve threat detection by using algorithms that learn from previous incidents and can adapt to new attack methods.

- regulatory compliance. With the increasing number of regulatory requirements for data protection, using AI can help organizations effectively manage risks and comply with the necessary standards.

- reducing costs. Implementing AI in information security processes can reduce security costs by increasing efficiency and reducing the number of false positives.

- security innovation. AI opens up new opportunities for developing innovative solutions in the field of data protection, such as biometric authentication, intelligent access control systems, and others.

The purpose of the study is to analyze the possibilities of using AI technologies applied in the field of information security, to study the problems in this area and to assess the prospects for using artificial intelligence to improve the level of information security of organizations.

2. Modern threats to information security of organizations

The number of incidents in 2023 and 2024 (Fig. 1), data on which are presented in the study by Positive Technologies (PT) [4], testifies to the global threats to information security.

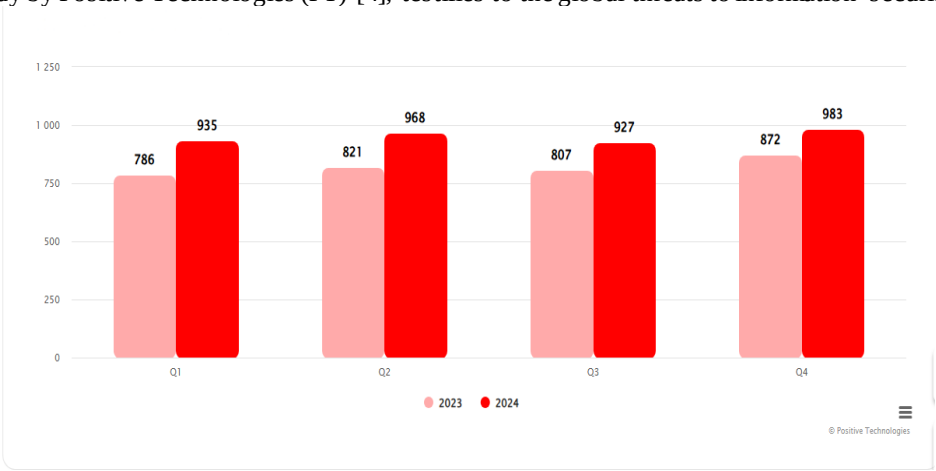


Fig 1. Number of incidents in 2023 and 2024 (by quarter)

According to PT Expert Security Center [4], in the fourth quarter of 2024, the number of incidents increased by 5% compared to the previous quarter and by 13% compared to the same period last year. Malicious software (malware) is the main tool of attackers: it was used in 66% of successful attacks on organizations and 51% on individuals. Ransomware (42%) and remote-control malware (38%) were most often used against organizations, while spyware (48%) was used against individuals. There was an increase in interest in spyware in attacks on organizations (20%, 4 percentage points more than in the previous quarter). During the period under review, confidential information was disclosed as a result of 53% of successful attacks on organizations, and disruption of the organization's core business was observed in 32% of incidents.

Figure 2 provides information on the industry affiliation of organizations that are victims of cyberattacks in 2024 [4]. It should be noted that 3% of successful attacks were directed at transport organizations [5].

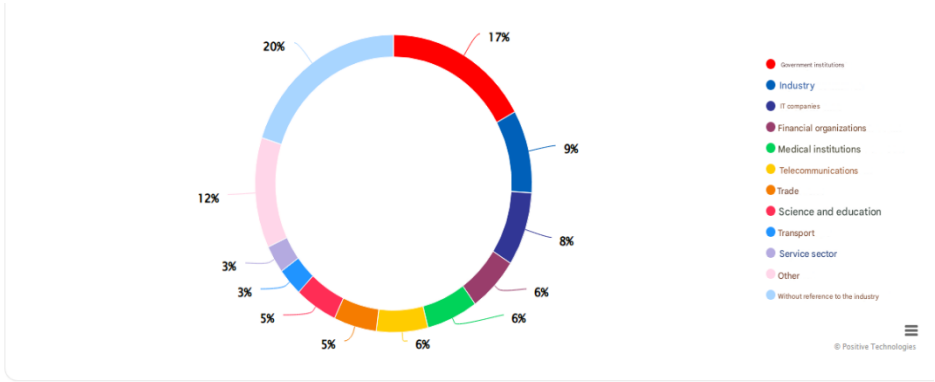


Fig. 2. Organizations that are victims of cyberattacks in 2024

Data on the malware tools used in hackers' arsenals in successful attacks on organizations and individuals in the fourth quarter of 2024 are shown in Figure 3 [4].

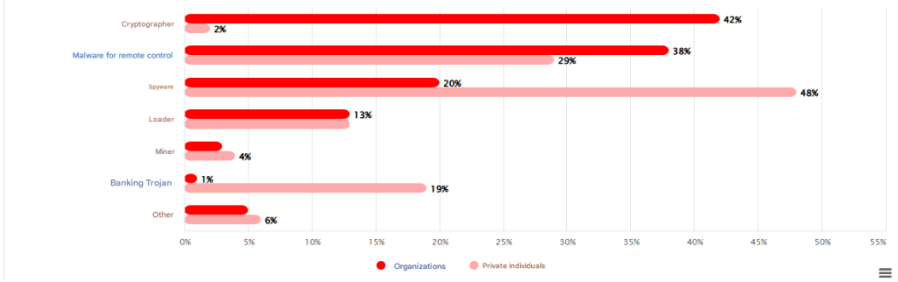


Fig. 3. Share of successful attacks using malware in the fourth quarter of 2024

In the fourth quarter of 2024, ransomware (42%), remote control malware (38%) and spyware (20%) were most often used in successful attacks on organizations. Ransomware continued to be a relevant threat to organizations around the world and was actively used not only for financial gain, but also to cause damage for hacktivist purposes.

In successful attacks on individuals in the fourth quarter of 2024, the trend towards using spyware (48%) and an increase in interest in banking Trojans (19%, 8 percentage points more than in the previous quarter) was observed.

The share of remote access malware in successful attacks on organizations and individuals in the fourth quarter of 2024 decreased by 6 percentage points compared to the previous quarter and amounted to 38% and 29%, respectively [4].

3. Possibilities of using artificial intelligence technologies in ensuring the information security of organizations

Let us consider the main tasks of using AI in ensuring information security, as well as defense tactics and technologies that use artificial intelligence.

AI technologies make it possible to solve many problems at all stages of preparation for repelling an attack, as well as during detection, response, and liquidation of the consequences of an incident. Each application can be attributed to one of three main tasks: reducing the workload of cybersecurity specialists, detecting anomalies in the behavior of

users, applications, and systems, advanced threat detection, and automation of defense systems [6]. Table 1 presents the main tasks of AI technologies in ensuring the information security of organizations and their content [7].

Table 1. Main tasks of AI technologies in ensuring information security of organizations and their content

Tasks	Contents
1. Reducing the workload and helping specialists	Using AI technologies in security solutions, it is possible to automate routine processes of primary processing of security events and other information that specialists are currently forced to analyze. In addition, chatbots based on large language models (LLM) can provide operational support to specialists in the decision-making process to counter cyber threats.
2. Threat and incident detection	AI technologies, processing large volumes of data, allow identifying malicious actions, non-standard behavior of users and systems, and providing early warning of incidents. In this case, such data as logs are used - files that record events of network equipment and servers, telemetry of end devices, network traffic, authentication data, and others. The following methods are used to detect anomalies and advanced threat detection: machine learning and statistical analysis, graph methods, sequence analysis, anomaly detectors, and others.
3. Incident Investigation and Response Automation	AI-based solutions can automate not only attack detection processes, but also decision-making, response, and incident prevention. AI technologies can be used to speed up investigations, collect context, and automated response measures. For this purpose, such methods as automated playbooks, ML optimization of investigation routes, and incident categorization scenarios are used. In this way, automated or semi-automated measures (device isolation, account blocking, file quarantine) are implemented, the average response time is accelerated, the share of automated incidents and the percentage of closed incidents without manual intervention are increased.

Table 2 presents the capabilities of AI technologies at individual stages of information security [7].

Table 2. Capabilities of AI technologies at individual stages of information security

Stages	Capabilities
1. Warning and forecasting of threats	AI technologies can automatically detect attempts to unauthorized access to systems and networks, reducing incident response times. Machine learning algorithms can use historical and observational data to predict future attacks and vulnerabilities, allowing organizations to take proactive measures to protect themselves.
2 Vulnerability Management	AI can help assess software and system vulnerabilities, suggesting prioritized actions to address them. For vulnerability assessment, AI can automatically scan systems and applications for known

	vulnerabilities, using vulnerability databases and machine learning algorithms to identify potential weaknesses. For organization security analysis AI can analyze the configurations of an organization's systems and networks, identifying misconfigurations or deviations from best practices that could lead to vulnerabilities.
3 Penetration testing and attack simulation	Conducting penetration tests with AI technologies can automate the penetration testing process by simulating the actions of attackers and identifying vulnerabilities in the system that can be exploited for unauthorized access. AI can help in risk assessment by analyzing various factors, such as the likelihood of an attack and the potential consequences, which allows organizations to make more informed decisions about security measures. Generative AI can be used to create a set of the most likely passwords for a specific target, analyze text files of the system under test, and generate an understandable report on the results of the attack simulation.
4 OSINT (open-source intelligence) of current cyber threats	AI technologies can significantly improve open-source intelligence (OSINT) processes and help identify relevant cyber threats to ensure information security in the following ways: - automated data collection. AI can automate the process of collecting information from various open sources, such as social media, forums, blogs, news sites, and vulnerability databases. This allows for quick and efficient collection of relevant data on current threats; - big data analysis. AI can process and analyze large amounts of data, identifying patterns and trends that may indicate new cyber threats. This includes using machine learning algorithms to classify and predict threats; - anomaly detection. AI can identify anomalies in user or system behavior that may indicate potential threats. For example, a sharp increase in discussions about a certain vulnerability or attack may be a signal that deeper analysis is needed; - sentiment analysis. AI can conduct sentiment analysis of open-source texts to determine public opinion on specific threats or incidents. This can help in assessing the level of risk and community response to specific cyber threats; - threat geolocation. Using geographic data, AI can visualize the distribution of threats in different regions, which helps organizations better understand the risks based on their location; - mention monitoring. AI can track mentions of specific companies, technologies, or vulnerabilities in open sources, providing timely notifications about new threats or incidents; - social media analysis. AI can analyze social media data to identify potential threats related to cyberattacks, including identifying attackers and their modus operandi.
5 Code verification	AI technologies offer a variety of options for code review to ensure information security:

	- static code analysis. AI can be used for automated static code analysis, identifying vulnerabilities and potential errors in software before it is executed. Algorithms can learn from known vulnerabilities and patterns to find similar problems in new code; - dynamic analysis. AI can help in dynamically analyzing applications during their execution, monitoring the program's behavior and identifying anomalies that may indicate vulnerabilities or attacks; - natural language processing. AI can analyze code documentation and comments to identify potential problems or flaws in logic that can lead to vulnerabilities; - automated testing. AI can generate test cases based on code analysis, allowing for more efficient testing of application functionality and security; - learning by example. Machine learning algorithms can learn from existing code and security incident data to improve their predictions and find new types of vulnerabilities; - monitoring code changes. AI can monitor changes in the code base and automatically check new commits for potential security issues; - recommendations for fixes. Based on code analysis, AI can provide recommendations for fixing found vulnerabilities, making it easier for developers to eliminate them; - identifying code patterns. AI can identify code patterns that often lead to vulnerabilities and suggest best practices for avoiding them.
6 Control of confidential data	AI technologies can be useful for controlling confidential data in ensuring information security in the following areas: - classification and identification of sensitive data; - automated recognition of personal data, financial information, organizational secrets, etc. in texts, files and databases; - detection of attempts to transfer confidential information via channels, in the cloud, in messengers, to external media; - unusual access patterns, attempts to copy data, unusual hours of activity, combinations of users and resources; - risk-based access control and risk assessment of each access request, multi-factor and contextual authentication, dynamic access rights; - context-dependent use of encryption, replacement of confidential values with tokens, masking for developers and testing; - automated incident detection and generation of response actions; access control and security policies for large cloud services; - data traceability and data management, automatic tracking of data sources, transformations and usage for auditing and compliance with information security requirements; - automatic removal or replacement of sensitive elements in documents and messages. All these capabilities have the potential to protect against confidential data leakage, the most common (54%) consequence of cyberattacks on organizations in 2024 [8].
7 Detection of malicious and	The use of artificial intelligence in protecting organizational information systems involves the automatic analysis of large volumes

abnormal activity, unknown threats	of data (logs, network traffic, security events, end-device telemetry, etc.) to detect malicious and abnormal activity, speed up responses, and improve detection accuracy compared to traditional signature-based methods (classic virus detection methods based on comparing files with a database of known virus signatures). At this stage, AI solves the following information security problems: - detection of known and unknown malicious objects: malware, account compromises, C2 activity (Blocking C&C traffic or "dismantling" an attacker's C2 infrastructure can stop a cyberattack), zero-day exploits (a type of malware) through behavioral characteristics; - detection of anomalies and deviations: unusual user behavior patterns, sudden changes in traffic, authorization logs, access to data; - detection of unauthorized actions and abuses: suspicious financial transactions, phishing and malicious attachments, credential manipulation; - early warning and reduced response time: accelerated incident identification, automatic correlation analysis between different data sources; - decision support in SOC (Security Operation Center): assisting analysts with explainable conclusions, automatic context collection and response recommendations.
8 Handling security events	According to Microsoft [9], SOC teams receive an average of 4,484 alarms per day and spend about 3 hours manually separating real threats from the noise. At the same time, it takes a SOC employee about 10 minutes to process and analyze each event. In such conditions, false positives become a problem, which inefficiently spend the efforts and time of specialists, as a result of which they can miss a real attack. False positives occur due to the fact that some actions of intruders can be disguised as legitimate user activity. The use of artificial intelligence can potentially solve the problem of SOC overload. An AI solution will be able to conduct an initial sorting of security events, remove likely false positives and highlight incidents that really require attention for a specialist. The use of AI technologies will significantly (according to IBM estimates, twice) speed up the sorting and processing of security events, and therefore increase the overall efficiency of the SOC [9].
9 Access analytics and behavioral analysis (UEBA – User and Entity Behavior Analytics)	Behavioral analysis in information security is an approach that relies on modeling the normal behavior of users and other objects (devices, applications, services) and identifying deviations from it as potential threats. Its goal is to detect account compromise, insider threats, illegal data movement and other types of abuse earlier. To implement it, the behavior of: users (logins, time and place of login, attempts to access data, sequences of actions); objects (devices, servers, cloud services, applications, containers); networks and applications, workflows and anomalies in access to resources and data. Practical examples of such behavior may be: unusual login activity outside working hours from a new device or from a different geolocation, and then access to sensitive data; a step-by-step chain of

	actions: login → access to critical files → copy to external media or upload to the cloud; a sudden increase in outgoing network traffic from the server or non-standard requests to the cloud service API. Using this approach with machine learning technologies, it is possible to create a profile of the normal operation of an object, such as a user, system or network, from which deviations (anomalies) may indicate the actions of a cybercriminal. As a result, this can lead to early detection of anomalous behavior, insider threats and account compromises, reducing the time to detect threats and respond.
10 Recognition of phishing, unwanted content, dangerous sites	Social engineering remains one of the main methods of cybercriminals in 2024, it was used in every second attack on organizations. Moreover, in 42% of attacks using malware, the delivery channel was phishing emails [4]. Cybercriminals manipulate the emotions of victims to achieve their goal, so in protecting against phishing, employee training alone is not enough: anyone can make a mistake and not check a malicious email. AI can help at different levels of protection against phishing and dangerous sites - from URL and page content analysis to user behavior and network signals. Together, such solutions provide early warning, automatic blocking and clear explanations to the user. For this, machine learning technologies are used that implement URL analysis and page features (URL length, number of subdomains, use of suspicious domain suffixes, presence of words frequently used in phishing, domain age, ownership history, IP address) and allow determining the likelihood of risk and the threshold blocking signal. The use of such an approach can help reduce the risk of early phishing through rapid online risk assessment and automatic blocking; provide comprehensiveness (URL, content, visual authenticity and reputation in one solution); increase flexibility (can be adapted for web feeds, email (for attachments/links) and mobile serial traffic); the ability to learn from real users through feedback, improving accuracy over time.
11 Reaction	Modern AI technologies can significantly speed up incident response and increase organizational resilience through the following solutions: - early detection and accelerated investigation; - ML algorithms for UEBA and behavioral analytics detect anomalies in user and entity behavior, combine signals from different sources and reduce the number of false positives for incidents; - response with automatic rule application: node isolation, account blocking, firewall rule update, blocking malicious IP addresses; - automation of routine remediation actions: revert changes, change keys/passwords, stop malicious processes; - quick verification of backup integrity, automatic restoration of a safe version, audit of changes; - post-incident analysis based on ML: determining the causes of the violation, assessing the impact and steps to prevent recurrence. Such support allows specialists to obtain the necessary data about a security incident much faster, and therefore make a decision more quickly. One of the promising areas is training organizations' own LLM

	cybersecurity assistants, who, if cybercriminals were able to achieve the goals of a cyberattack, can potentially help cope with the consequences of the incident and suggest optimal actions to eliminate the damage caused.

An analysis of the information security capabilities considered in Table 2, including various solutions, allows us to divide them into three levels depending on whether they use artificial intelligence technologies: AI is already used, AI can be used, the use of AI is not justified or will not bring significant benefits.

Thus, the capabilities of AI in information security are quite broad, but even with further development of the technology, there are still unused areas of protection.

4. Problems of using AI technologies in ensuring information security of organizations

High expectations for the use of AI technologies in information security face a number of challenges related to the cost and complexity of their implementation: scarce and high-quality training data is required; only highly qualified specialists with certain skills and competencies can create such a product; the use of AI solutions requires significant computing power, which may be unavailable to organizations for various reasons and / or require significant support from the business.

There is also a risk of excessive automation: not every decision should be made automatically. In combination with the competent expertise of a human analyst, it is important to ensure the correct interpretation of the results.

It is necessary to consider the security of the model itself, protection against leakage of confidential information and prevention of abuse of generative AI models.

Regular audit of models is necessary for deviations and false positives. In addition, there are ethical and legal frameworks: attack modeling should only take place in authorized environments with clear restrictions and access control.

5. Prospects for using AI technologies in ensuring information security of organizations

Research in the field of information security highlights the following trends in the use of AI in this area [7,10]:

Autopilot. An autonomous system based on AI ("autopilot" in the context of information security) can improve the protection of an organization due to the following capabilities:

- rapid response to incidents: automatic detection, prioritization of threats and execution of predefined actions without delays;
- increased coverage and accuracy: continuous monitoring 24/7, reduction of omissions and errors due to human fatigue;
- standardization of response: uniform schemes for actions to automate response (playbooks) and approaches to incidents, minimization of variability of actions;
- effective risk management: early detection of abnormal behavior, vulnerabilities and configuration errors, predictive analytics on threats;
- resource savings: reduced load on SOC/IT departments, redistribution of people to more complex tasks;

- improved decision making: explain ability of decisions (when necessary) and transparent action logs for audit and compliance.

Cyber Range Modeling. A cyber range is an isolated, controlled environment that simulates real networks, applications, and data for training employees and testing defenses. Using AI technologies in such modeling allows for automatic creation of attack scenarios, generation of realistic traffic, adaptive adjustment of defenses to the evolution of threats, and acceleration of training and testing of information security measures.

AI improves the efficiency of cyber range modeling in the following ways:

- threat and attack scenario modeling. AI can automatically generate a variety of attack scenarios and their variations (different tactics, techniques, and procedures). Generative models can create new attack variants to test the resilience of defenses to previously unknown scenarios, within an ethical and controlled environment.

- generation of realistic data and traffic. Machine learning maintains the realism of the load: artificial but plausible traffic and user events. This allows testing SOC operations and detection methods in the absence of real incidents.

- detection, analysis and automatic response. Behavioral analytics models identify unusual patterns and early signs of attacks.

Within the cyber polygon, AI can help with incident prioritization, automatic early detection, and response management (automated isolation, blocking, and escalation scenarios). It can safely explore which security settings work against potential attackers. AI agents can act as an adaptive "red team," trying to bypass defenses, but within a safe environment and ethical constraints. This approach accelerates the search for weak points, testing of regulations, and increasing the resilience of response processes. In addition, AI optimization of training scenarios is possible: adjusting the complexity depending on the level of employee training, automated generation of reports after training to improve further actions.

6. Conclusion

Thus, the study of the possibilities of using AI in information security is relevant and necessary to improve the level of protection of organizations in the context of modern digital development of society.

Analysis of the possibilities of using AI technologies and methods applied in the field of information security helps to identify their potential and effectiveness, as well as the creation and testing of new algorithms and models based on AI to improve the level of protection of organizational information from cyber threats.

Despite a wide range of possible tasks for AI, a person in the field of cybersecurity is still irreplaceable. Even with a level of technology that is significantly superior to today, there will still be many tasks for a specialist, such as developing a general defense strategy, monitoring and managing the operation of AI tools, and solving complex and non-standard issues. In the future, the ability to effectively work with AI tools will become an important component of the necessary competencies of an information security specialist.

References

1. Pugacheva, Olga. The use of Artificial Intelligence in business and society: threats and regulation "Economic Security in the Context of Systemic Transformations", international conference (3; 2023; Chişinău). Proceedings of International Conference "Economic Security in the Context of Systemic Transformations", 3rd Edition, December 7-8 2023, Chişinău / drafting committee: Tatiana Bucos [et al.]. – Chişinău : SEP ASEM, 2024. – 271 p. (p. 249-259)

2. Pugacheva, O. V. Use of artificial intelligence in business. Actual issues of modern economic science: theory and practice: collection of scientific articles. Issue 2. Gomel: GSU named after F. Skorina, 2023, pp. 209-212.
3. Pugacheva, O. V. Use of artificial intelligence in the economy and society: directions, problems and regulation. // Proceedings of Francisk Skorina Gomel State University, 2024, 5 (146), pp. 138-156.
4. Current Cyber Threats: IV quarter of 2024 – I quarter of 2025. https://telematika.com/press/news/kak_iskusstvennyy_intellekt_menyaet_pravila_kiberbezopasnosti_itogi_diskussii_na_trans_ai_2025/.
5. How Artificial Intelligence is Changing the Rules of Cybersecurity: Results of the Discussion at Trans AI 2025. https://telematika.com/press/news/kak_iskusstvennyy_intellekt_menyaet_pravila_kiberbezopasnosti_itogi_diskussii_na_trans_ai_2025/
6. Trends in Malware Development and Its Role in Cyberattacks. https://www.ptsecurity.com/ru-ru/research/analytics/trendy-v-razvitii-vredonosnogo-po-i-ego-rol-v-kiberatakah/#id13
7. Artificial Intelligence in cyber defense. https://www.ptsecurity.com/ru-ru/research/analytics/iskusstvennyi-intellekt-v-kiberzaschite/#id31
8. Cybercrime Market. https://www.ptsecurity.com/ru-ru/research/analytics/rynok-kiberprestupnosti/#id
9. Cybersecurity incident correlation in the unified security operations platform. https://techcommunity.microsoft.com/blog/microsoftthreatprotectionblog/cybersecurity-incident-correlation-in-the-unified-security-operations-platform/4214394
10. Artificial Intelligence in Cyberattacks. https://www.ptsecurity.com/ru-ru/research/analytics/iskusstvennyj-intellekt-v-kiberatakah/#id1

NVivo Application for Qualitative Data Analysis in Scientific Research: Options and Result Visualizations

Tatiana Colesnicova, Mihail Ciobanu

Academy of Economic Studies of Moldova, Chisinau, Moldova
ctania@gmail.com, ciobanu.mihail.s@gmail.com

Abstract. The paper analyzes how NVivo application uses advanced tools for data organization, thematic analysis, and graphical representation of research findings from qualitative studies in the social and economic sciences. A special attention is given to NVivo's charts, models and word maps that supports the interpretation and communication of research findings. The research concludes that NVivo represents a valuable resource for increasing analytical quality of scientific work and effectively presenting complex qualitative results of research findings. The article was elaborated within the Project selected in competition "Stimulating excellence in scientific research 2025-2026" 25.80012.0807.37SE "The perfection of instruments for reducing income inequality and reducing poverty of the population in the process of European integration of the Republic of Moldova".

Keywords. software, application, data, analysis, visualization.

1. Introduction

Qualitative research in social and economic sciences often generates large volumes of unstructured data – interview transcripts, focus group discussions, open-ended survey responses, field notes, etc. Analyzing such data manually can be laborious and prone to bias. NVivo is a Computer-Assisted Qualitative Data Analysis Software (CAQDAS) designed to help researchers organize, code, analyze, and interpret these unstructured datasets.

Before tools like NVivo, qualitative analysis relied on physical methods (e.g. highlighters, scissors, index cards) to mark and sort themes, which was time-consuming and made it difficult to manage complex projects. NVivo provides a centralized digital workspace where all data sources (text, audio, video, PDFs, images) can be imported and managed within a single project. This centralization and organization allow researchers to handle dozens or even hundreds of documents systematically, maintaining a clear audit trail of the analysis process. NVivo became especially popular in academic research, government and policy studies, and lengthy interview-based projects, filling a gap for more rigorous and efficient qualitative analysis tools.

Over the past decades, NVivo has evolved into a leading QDA software. It not only supports basic tasks of coding and retrieving text segments, but also offers advanced features for linking data, shaping ideas, and even modelling relationships between concepts. The use of NVivo and similar CAQDAS has been shown to enhance the quality and systematic nature of qualitative analysis. By automating routine tasks (like grouping excerpts by code or searching for keywords), NVivo frees up researchers' time to explore deeper patterns and interpretations in the data.

Ultimately, tools like NVivo aim to make qualitative analysis more rigorous without sacrificing the richness and context that are the hallmarks of qualitative inquiry. This article provides an overview of NVivo's capabilities in qualitative data analysis, with special attention

to its options for data organization, thematic coding, and the visualization of results through charts, models, and "word maps" (word-based visualizations).

2. Literature review

The academic literature on qualitative methodology has increasingly acknowledged the value of CAQDAS tools like NVivo. NVivo is one of the most widely used QDA software packages and has been cited as the de facto choice for many qualitative researchers due to its extensive features and proven utility (QSR International, 2026). Early qualitative analysis software (such as NUD*IST in the 1980s) laid the groundwork for NVivo, which was developed as an improved successor with enhanced capabilities (e.g. allowing rich text handling, multimedia support, and collaborative work). NVivo's design is intentionally methodology-agnostic, meaning it can be applied to a range of qualitative approaches including grounded theory, phenomenology, discourse analysis, ethnography, content analysis, and mixed-methods research. This flexibility has made NVivo applicable in diverse fields from health care and education to sociology and economics.

Multiple sources have noted that using NVivo can increase the efficiency and quality of qualitative analysis (Bazeley & Jackson, 2013) (Silver & Lewins, 2014). Maher et al. (2018:1) highlight that NVivo "provide[s] excellent data management and retrieval facilities that support analysis and write-up" by easing the manual labour of coding and sorting data. The software replaces tedious manual tasks (cutting and color-coding piles of transcripts) with more efficient digital operations, allowing the researcher to focus on interpreting meaning rather than on data management mechanics. As a result, the analysis process becomes more systematic and transparent. NVivo maintains a searchable database of all coded excerpts, memos, and links, which contributes to the auditability of qualitative research – an important criterion for rigor. Additionally, NVivo supports teamwork by providing project merge tools and collaboration features. Researchers working in separate NVivo project files can merge their coded data into one master project, making NVivo ideal for team-based qualitative projects where multiple analysts are involved. This capability helps ensure consistency in coding and allows cross-validation of findings among team members.

Despite its benefits, the literature also discusses challenges in using NVivo. A commonly cited issue is the steep learning curve. Mastering NVivo's interface and features often requires formal training or significant practice, which can be a barrier for new users or those outside academia. NVivo is fundamentally a powerful data management and facilitation tool, but the intellectual work of analysis – identifying themes, interpreting patterns, drawing conclusions – remains with the researcher. As Zamawe (2015:13) puts it, "unlike statistical software, the main function of CAQDAS is not to analyse data but rather to aid the analysis process, which the researcher must always remain in control of". This is a critical point to dispel the misconception that qualitative analysis can be fully automated. Researchers must still apply their theoretical lens and contextual understanding. NVivo simply provides an organized framework to do so more efficiently and systematically. Some authors also warn about potential overreliance on code frequency or query outputs – just because a theme is mentioned frequently (as NVivo's queries might show) does not automatically mean it is the most important insight in the study. Qualitative analysis requires nuance and considering context, something NVivo outputs cannot determine on their own (Knaflic, 2015).

Numerous published studies have documented using NVivo to strengthen qualitative analyses. For instance, in policy and program evaluation research, NVivo has been used to manage large sets of interviews and focus group data, making it easier to trace how conclusions were derived from raw data. Health researchers have utilized NVivo to maintain rigor in literature reviews and meta-syntheses by systematically coding findings from many sources.

In dissertation and thesis work NVivo help students handle extensive qualitative datasets and maintain consistency over long research projects. NVivo's role in mixed-methods research is to combine qualitative interview analysis with quantitative survey data by using classification and query features to compare groups or to integrate quantitative attributes (like demographic data) with qualitative themes. Overall, the literature suggests that when used appropriately, NVivo can improve the depth and transparency of qualitative analysis, though it requires careful methodological consideration and researcher skill to fully capitalize on its capabilities.

3. Methodology

This article employs a descriptive and exploratory approach to examine NVivo's options for qualitative data analysis and result visualization. It was conducted a review of NVivo's software documentation, user guides, and relevant methodological literature to outline the key features of the application. The focus is on NVivo 12 and later (including NVivo 15/Lumivero) as used in typical social science research projects. In addition, it was drawn on example use cases from existing research to illustrate how NVivo's tools are applied in practice.

For the purposes of illustration, it can be considered a hypothetical qualitative study on economic inequality in a community – involving data from interviews, focus groups, and field observations. Using such a scenario, we can walk through the stages of using NVivo to manage and analyze the data. All qualitative data (transcribed interviews, focus group transcripts, observation notes, relevant policy documents, etc.) are imported into NVivo as Sources. NVivo supports importing a wide variety of formats including text documents, PDFs, audio/video files (with or without transcripts), spreadsheets, and even social media content. Once in NVivo, these sources can be organized into folders or sets (e.g. grouping all interviews in one folder, focus groups in another). Researchers can also create case nodes for each interviewee or participant and assign demographic attributes (age, gender, etc.) to those cases. This structure allows for later comparisons across different participant groups. The organizational capability of NVivo ensures that no data get lost and that one can easily navigate between raw data and emerging themes. Coding is at the heart of qualitative analysis, and NVivo's interface is built to streamline the coding process. Users create nodes in NVivo which represent codes, themes, or concepts. These nodes can be arranged in a hierarchical tree (parent themes and sub-themes) or kept as free nodes for open coding. To code data, one simply highlights text (or time segments in audio/video) and assigns it to one or more nodes. NVivo's coding tools allow for both deductive coding (applying a predefined code framework) and inductive coding (creating new codes as patterns emerge from the data). The software tracks all coding actions – a piece of text coded at a node can be retrieved by opening that node to see all references in one place. This makes it easy to gather all evidence for a particular theme when it comes to analysis and writing. Memos can be written alongside to capture analytic insights or reflexive notes during coding. NVivo's ability to handle overlapping codes and to merge or reorganize nodes on the fly gives researchers flexibility as their understanding evolves. The end result of coding is a structured categorization of the qualitative data, which is the basis for deeper analysis. NVivo offers powerful query tools that enable asking complex questions of the coded data. For example, a Word Frequency Query can list the most common words or phrases in the text, which might hint at prevalent topics. A Text Search Query can find all instances of a particular term or phrase across the dataset (useful for checking how often certain concepts appear and in what context). More advanced queries include the Coding Query (to retrieve content coded at combinations of nodes, e.g. where "economic policy" and "education" co-occur) and the Matrix Coding Query, which can cross-tabulate two sets of categories (for instance, comparing themes across male vs. female participants, or across different focus groups). These query

results can be viewed in tabular form or exported. Such exploration helps in discovering patterns that might not be immediately obvious – for instance, one might find that certain themes are mentioned predominantly by a particular demographic group, or that some concepts frequently appear together, indicating a potential relationship worth investigating. By iteratively querying the data, researchers refine their insights and ensure that their interpretations are grounded in evidence from the dataset. If multiple researchers are coding data, NVivo includes tools for comparing coding consistency. The Coding Comparison Query can measure inter-coder reliability by checking the agreement on coding between different team members. In team settings, NVivo's Collaboration Cloud or server options enable multiple coders to work on the same project concurrently, or to periodically merge their work, as mentioned earlier. This methodological step is important for enhancing the credibility of the analysis in scientific research – high agreement on coding indicates that the themes are not just subjective to one analyst but are consistently observed. Any discrepancies can be discussed and resolved to refine the codebook. This methodology (data organization → coding → querying → team validation) sets the stage for producing meaningful results. With the dataset thoroughly coded and examined, we can now leverage NVivo's visualization features to represent the findings in a clear and compelling way.

4. Results

In the ***Table 1*** it is highlighted that Computer-Assisted Qualitative Data Analysis Software (CAQDAS) differs not only in technical capabilities but also in the research philosophy they best support. NVivo, ATLAS.ti, and MAXQDA emerge as comprehensive platforms suited for large, complex projects requiring deep coding, visualization, and mixed-methods integration. These tools are particularly appropriate for institutional research, doctoral work, and projects involving extensive datasets and advanced querying. In contrast, tools such as QDA Miner, Dedoose, and Quirkos prioritize accessibility, intuitive interfaces, and ease of learning, making them suitable for small-scale studies, teaching contexts, or teams requiring rapid onboarding. Their limitations appear mainly in advanced querying power and scalability for large multimedia datasets.

Open-source solutions (KH Coder, RQDA, Taguette, and Casandra) provide cost-effective alternatives, especially for researchers with technical skills or budget constraints. However, they often require greater methodological and technical expertise and may lack the polished interfaces and multimedia integration found in commercial platforms. Overall, the choice of CAQDAS should be guided not by popularity but by the nature of the research design, team capacity, data complexity, and the balance between analytical depth and usability.

Returning to NVivo software, after applying it to analyze the qualitative data, the findings can be organized around major themes and supported by visual representations. NVivo's advanced visualization tools – charts, models, and word-based maps – prove especially useful in interpreting and communicating the results of qualitative analysis. These visualizations translate the coded data and queries into accessible graphics, helping to summarize complex information and reveal patterns at a glance.

Using NVivo, can be identified several overarching themes, for example, regarding income inequality. For instance, common themes might include "Employment" and "Income", "Education" and "Skills", "Health", "Policy" and "Support". Each of these can have multiple sub-themes (e.g., "informal employment", "access to training programs", "healthcare access," etc.), which are organized in NVivo as hierarchical nodes. The coding process results in a tree of nodes reflecting the complexity of the issue. NVivo's coding stripes and node frequency counts shows which themes are most referenced across interviews. For example, "Employment" may appear in nearly every interview, whereas a theme like "Digital" may be

mentioned in only a few focus groups, indicating it is a less universal but still notable issue. This frequency information is automatically tallied by NVivo and can be visualized.

Table 1. Comparative overview of major CAQDAS (Computer-Assisted Qualitative Data Analysis Software) tools: strengths, limitations, and recommended use cases

Software	Strengths	Limitations	Use
NVivo	Handles text, audio, video, images, social media; strong querying (Boolean, matrix coding); rich visualizations; large user community	Expensive; steep learning curve; Mac version lags; project merging can be tricky	Comprehensive qualitative projects needing deep coding & visualization
ATLAS.ti	User-friendly interface; strong network visualizations; robust coding; multimedia support; good collaboration tools	High cost; some advanced queries less powerful than NVivo	Researchers who want visual, intuitive exploration of qualitative data
MAXQDA	Supports qualitative + mixed-methods; survey & statistical integration; strong longitudinal analysis; intuitive interface	Full version expensive; some advanced queries less deep than NVivo	Mixed-methods research, combining qualitative and quantitative
QDA Miner + WordStat	Strong text + quantitative analysis; statistical content analysis; integrates with surveys	Less focus on multimedia; user interface less modern	Mixed-methods, content analysis with quantitative/statistical needs
Dedoose	Web-based; cross-platform; good for distributed teams; balances ease and features	Internet-dependent; privacy concerns; less suited for huge datasets	Collaborative teams, cross-platform qualitative analysis
Quirkos	Very simple, intuitive, visual interface; fast learning curve	Limited advanced querying; weaker on large projects or multimedia	Small-scale projects, beginners, teaching
KH Coder (open source)	Text mining, clustering, co-occurrence analysis; free	More technical; limited multimedia; steep learning for non-coders	Large text corpora, researchers with coding/statistical background
RQDA (open source, R-based)	Free; integrates with R for powerful workflows	Limited interface; R knowledge required; development not always active	Researchers comfortable in R, reproducible coding
Taguette (open source)	Free, simple, easy for text-only coding	Few advanced features; text only	Budget-limited projects, teaching, simple coding
Cassandre (open source)	Collaborative (Cassandre); free	Interfaces less polished; fewer updates	Budget-conscious teams, experimental or academic use

Source: elaborated by authors

One of NVivo's classic visualization tools is the chart view. A chart of coding references by theme can be generated to see the distribution of data across the major categories. NVivo allows users to create bar charts or pie charts based on coding frequency or

coverage for selected nodes. For example, a bar chart illustrated, for instance, that "Employment" and "Support" had the highest number of references (indicating its prominence in discussions), followed by "Digital" and "Access", and with "Capacity" being the least referenced theme. This chart provides a quick quantitative snapshot of qualitative findings – a helpful way to validate that the analysis covered all major areas and to communicate which topics were most salient to participants (it is important to note, however, that frequency is not the sole indicator of importance; even less frequent themes can be critical, which is why they are interpreted within context).

To get a sense of the language and emphasis used by participants, we conducted a word frequency query on some interview transcripts. The query results were used to create a word cloud (sometimes called a "word map"). Figure 1 shows an example of an NVivo word cloud generated from qualitative text. In a word cloud, the size of each word corresponds to how often it appears in the corpus.

In our analysis, the word cloud highlighted prominent terms such as "employment", "digital", "support", and "access", suggesting these concepts were central in participants' discussions. NVivo automatically filters out common stop-words (like "and", "the", "is") unless configured otherwise. Using word clouds in qualitative research helps in highlighting the prominent terms and key themes in research, simplifying the presentation of core findings. We found this to be true: the word cloud provided an intuitive visual summary of what issues were most frequently talked about, which was a helpful starting point for deeper analysis and for explaining results to stakeholders who might not read the full report. It served as a conversation piece to quickly convey the focus of discussions before delving into detailed results.

NVivo also enables researchers to visually map relationships between concepts using models or maps. We utilized NVivo's modeling tools to create a concept map of how different factors related to income inequality interact. Figure 2 illustrates a simplified example of an NVivo concept model (network diagram) showing key concepts and their interconnections.

In our result, the model revealed, for instance, that Education Level was linked to Employment Opportunities, which in turn was connected to Income Stability. We also linked Policy Support as an overarching factor that influences both education and jobs (for example, government training programs affecting employment prospects). NVivo's models are essentially mind maps or concept maps created within the software's "Model Editor", allowing the researcher to drag and drop nodes (which can represent themes, cases, or any project item) and draw connections or arrows to denote relationships. This visual representation supported our analysis by making implicit relationships explicit. Models in NVivo map connections between concepts, uncovering relationships that might not be apparent in text-based analysis. In other words, by arranging concepts visually, we could better see the network of influences in a way that a linear text list of themes might not reveal. The model serves as a hypothesis of how things are connected, which we can then compare against the data (e.g., checking if participants who discussed "Education" also mentioned „Employment", etc., which NVivo can facilitate through query cross-tabs).

In addition to word clouds and concept maps, NVivo provides several other visualization options that can be leveraged for a comprehensive view of results. For example, we used a hierarchy chart (tree map) to visualize the relative size of sub-themes under each major theme, based on number of references. This quickly showed which sub-topics were most discussed within a theme. We also tried NVivo's cluster analysis on the interviews to see if any interviews grouped together based on thematic similarity – NVivo's cluster dendrogram suggested patterns (for instance, perhaps interviews from rural participants clustered separately from urban participants because of different issue emphasis). While such

exploratory visuals need cautious interpretation, they added depth to our results by suggesting patterns for further exploration. Furthermore, NVivo's matrix coding query results were exported as a table to show how often each theme appeared in each demographic group, facilitating a clear presentation of comparative results (e.g., differences between age groups in which themes they mentioned most).



Figure 1. NVivo word cloud of key themes identified in interview transcripts
Source: Elaborated by authors
Note: Frequently occurring terms from the qualitative data are displayed with larger font size, visually highlighting central topics such as "employment", "access", "digital", "support", "services", and "literacy" as reflected in participants' discussions.

Through NVivo, our qualitative analysis yielded a nuanced understanding of income inequality drivers. Employment opportunities and access to quality education emerged as the most frequently discussed factors in improving one's economic situation (visually evident from word frequency and coding charts). Health issues and access to healthcare were also common themes, linking poverty with health outcomes.

The concept model demonstrated that these issues are interlinked – for example, education is tied to employment, which in turn affects health and income, suggesting a cycle that can trap families in poverty without multi-faceted intervention.

Using NVivo's cases and classification, we observed some differences among subgroups. Younger participants stressed education and skills training more, whereas older participants focused on healthcare and pensions. Urban vs. rural differences were noted in terms of access to services (rural interviews highlighted transportation and infrastructure issues). These patterns were identified through coding queries and visualized via comparison diagrams.

A recurring theme was the role of government or NGO programs. Many participants referenced policy support – some noted positive impacts of certain social programs, while others pointed out gaps. This theme, while not the most frequent, was critical; we gave it special attention in our model to show it as an external influence on all other factors.

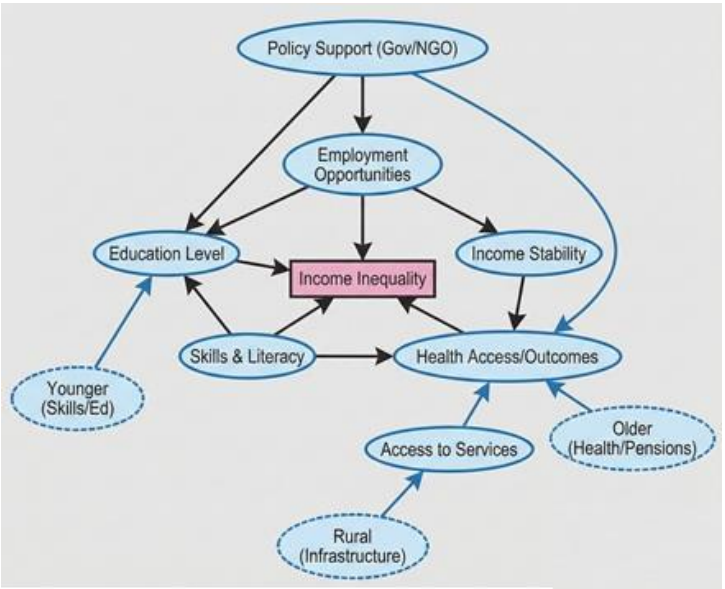


Figure 2. NVivo concept model of the interconnections between key themes related to income inequality
Source: Elaborated by authors
Note: The diagram illustrates how "Policy Support (Gov/NGO)" influences "Employment Opportunities", which in turn affect "Education Level" and "Income Stability" and contribute to "Income Inequality". The model also shows how "Skills & Literacy", "Health Access/Outcomes", and "Access to Services" interact within this system, while contextual factors such as "Young (Skills)", "Older (Health/Pensions)" and "Rural (Infrastructure)" shape participants' experiences. Arrows indicate the direction of influence between concepts as identified through qualitative analysis.

5. Discussions

All the findings are enriched and validated through NVivo's tools. The software ensures that every conclusion drawn could be traced back to multiple raw data excerpts, enhancing the credibility of the results. The results illustrate not only the substantive findings about the example topic but also how NVivo's features enhance the analytical process and presentation of qualitative research. One of the clear benefits observed is that NVivo allows researchers to engage with their data more deeply and systematically. By using tools like queries and models, with NVivo can uncover patterns and relationships that may have been missed with a purely manual analysis. For instance, the ability to swiftly retrieve all instances of a concept across dozens of documents or to visualize co-occurring themes pushes the researcher to consider connections in the data that are not immediately obvious. This aligns with the view that integrating data visualization into qualitative analysis can help "challenge conventional approaches" and encourage researchers to rethink assumptions. A concept map is a case in point – creating a visual model of how themes relate forced us to explicitly theorize the interplay between factors, which strengthened the analysis. Additionally, NVivo's mixed-methods capacity (e.g., allowing quantitative attributes and qualitative data to be linked) gives a richer perspective. It can, for example, quantitatively show variations in theme prevalence by group, adding a layer of empirical rigor to qualitative findings.

A major challenge in qualitative research is conveying dense, text-based findings to stakeholders (e.g., policymakers, community leaders, or even other researchers) in an accessible way. NVivo's result visualizations – charts, word clouds, models – directly address this challenge. They act as bridges between the qualitative evidence and the audience's understanding. Showing a word cloud to community partners can provide an immediate sense of the main concerns of participants without requiring them to read through pages of quotes. The visual impact of seeing keywords in large bold letters can resonate more quickly than just stating "employment and health are common themes". Likewise, the bar chart of theme frequencies and the concept map offers talking points that makes the presentation of results more engaging and clearer. NVivo's visualization tools can improve the uptake of research findings, as they cater to visual learning preferences and can summarize complex data efficiently.

Using NVivo also has benefits for the interpretative phase of the research. By having all data and coding centralized, we could easily cross-check interpretations. If a surprising pattern emerged (for example, an unexpected cluster of participants linking two ideas), we could click on the visualization (in NVivo, many charts and models allow direct clicking through to the underlying data) to see the actual excerpts behind it. This can ensure the discussion of results remained grounded. It also can help in maintaining context – one of the criticisms of data visualization in qualitative work is the risk of oversimplification. This can be mitigated by using NVivo's capability to attach annotations and memos to visual outputs. For instance, when presenting the word cloud, this can be accompanied with quotes for the top terms to preserve the original voices. NVivo makes it easy to pull those illustrative quotes directly from the nodes associated with each theme.

In discussing NVivo's role, it's important to acknowledge practical limitations we encountered. Learning to effectively use the software required an upfront investment of time – setting up the project, creating a coding schema, and learning the visualization functions. This is consistent with reports that NVivo is not intuitive for new users and often demands training. This is managed by consulting NVivo's help resources and starting with NVivo's sample project for practice. Another consideration is that while NVivo's built-in visuals are great for analysis and internal use, they may need polishing for publication or formal presentations. The bar charts and models, for example, are exported and then slightly refined (increasing font sizes, adjusting colours). Experienced researchers often export NVivo visuals to software like PowerPoint or Adobe Illustrator for final touch-ups. This is not a flaw per se, but users should be aware that NVivo's focus is on analysis; presentation-ready graphics might require additional formatting. Despite these points, the time invested in learning NVivo was recompensed by the efficiency and clarity it brought to our project.

Had we conducted this qualitative analysis without NVivo, it would have been considerably more cumbersome to achieve the same depth. Manual coding of dozens of interviews would risk errors and make it difficult to reorganize themes dynamically. Creating visualizations like concept maps or maintaining counts of themes manually would be prone to mistakes. NVivo essentially acted as an analyst's assistant – handling data management tasks swiftly and accurately, so that we could focus on interpreting what the data meant. It is worth noting, however, that the researcher's critical thinking remains the driver of analysis. NVivo did not tell us what the themes are or why patterns exist; it enabled us to systematically arrive at those insights ourselves. This reflects the consensus in methodology literature that CAQDAS tools support but do not replace analytical judgement. The meaningful explanations (for example, why education links to income, or how policy affects outcomes) comes from our understanding and theoretical grounding, enriched by NVivo's efficient data handling.

In summary, the use of NVivo in research increases the analytical quality and helps in effectively presenting complex qualitative findings. The thematic maps and visual outputs makes analysis more transparent and communicable (the story in the data can be seen and not just read). These advantages align with broader observations that NVivo and similar software, when used thoughtfully, can raise the standard of qualitative research reporting. Nonetheless, the advice is to approach NVivo as a powerful toolkit – one that requires methodological care in its use – rather than a push-button solution. As with any tool, its value is maximized when combined with the researcher's expertise and critical insight.

6. Conclusions

The exploration of NVivo's application in qualitative data analysis demonstrates that it is a valuable resource for enhancing both the analytical rigor and the presentation of findings in scientific research. NVivo's advanced tools for data organization, thematic coding, querying, and visualization enable researchers in social and economic sciences to handle complex qualitative datasets with greater efficiency and depth. By leveraging NVivo, analysis that once might have taken months of manual work can be completed more systematically – yielding results that are not only robust but also easier to verify and update as new data comes in. The software contributes to making qualitative analysis more transparent and auditable, as every code, link, and conclusion can be traced back to source data.

Crucially, NVivo's visualization capabilities (charts, models, word clouds, etc.) support the interpretation and communication of research findings in ways that traditional text-based reports often cannot. They bridge the gap between rich qualitative insights and clear storytelling: helping highlight key themes at a glance, illustrating relationships between ideas, and engaging diverse audiences. A well-crafted word cloud or concept map can distill complex narratives into an accessible visual form, sparking conversation and understanding among stakeholders. This resonates with recent perspectives that data visualization in qualitative research is more than just aesthetic – it is a means of translating knowledge for broader impact. NVivo equips researchers with the tools to create such visuals within the analysis process itself, ensuring that visualization is integrated with, not separate from, analytic thinking.

Of course, using NVivo (or any CAQDAS) is not without challenges. Researchers must invest time in learning the software and be mindful to preserve the context and richness of the data when presenting visual summaries. However, as our discussion noted, these challenges can be managed through careful planning, training, and thoughtful use of the software's features. The benefits – in terms of enhanced insight and effective communication – tend to far outweigh the initial hurdles. It is also important to reiterate that NVivo aids but does not replace the researcher's role. Insightful qualitative analysis still depends on the researcher's critical interpretation, theoretical knowledge, and creativity in probing the data. NVivo simply provides a powerful platform to support this work, allowing researchers to go beyond surface-level analysis and uncover deeper stories within their data.

NVivo represents a significant step forward in the toolkit of qualitative researchers. By improving how data is organized, analyzed, and visualized, NVivo contributes to more rigorous and impactful qualitative research. For studies dealing with complex social and economic issues – where understanding nuance and conveying complexity are both essential – NVivo offers options that enhance analytical quality and enable the effective presentation of findings. As qualitative research continues to grow in prominence across disciplines, tools like NVivo will play an important role in ensuring that qualitative insights are systematically derived and powerfully communicated, ultimately bridging the gap between rich data and real-world decision-making.

References

1. Bazeley, P., & Jackson, K. (2013). Qualitative data analysis with NVivo. Sage Publications.
2. Maher, C., Hadfield, M., Hutchings, M., & de Eyto, A. (2018). Ensuring rigor in qualitative data analysis: A design research approach to coding combining NVivo with traditional material methods. International Journal of Qualitative Methods, 17, 1–13. https://doi.org/10.1177/1609406918786362
3. QSR International. (2026). NVivo (Version 15) [Computer software]. Lumivero. https://www.qsrinternational.com/nvivo-qualitative-data-analysis-software
4. Silver, C., & Lewins, A. (2014). Using software in qualitative research: A step-by-step guide. Sage Publications.
5. Zamawe, F. (2015). The implication of using NVivo software in qualitative data analysis: Evidence-based reflections. Malawi Medical Journal, 27(1), 13–15. https://doi.org/10.4314/mmj.v27i1.4
6. Knaflic, C. N. (2015). Storytelling with data: A data visualization guide for business professionals. Wiley. ISBN: 978-1-119-00225-3

Constructing a Conceptual Model for the Modernization of a Small Family-Owned Enterprise

Ivan Stamov

Academy of Economic Studies of Moldova, Chisinau, Republic of Moldova
ivanstamov1977@gmail.com

Abstract. This study presents a conceptual model for the modernization of a small family winery, substantiated through a single case study. In contrast to linear approaches focused primarily on investment, the model demonstrates that sustainable development is achieved through the synchronous development of interrelated elements — or "nodes": technological, cognitive, and capital. Based on interviews with founders, analysis of financial statements, and a survey of several small wineries, systemic barriers hindering the development of the majority of small producers are identified. The key conclusion emphasizes the need for institutional support for small wineries through the implementation and expansion of the protected geographical indication system (IGP/DOP), which synchronizes the development of all model elements and ensures the market capitalization of quality.

Keywords: Family business, small wineries, technological modernization, conceptual model, knowledge management, economic sustainability, IGP, Moldova.

1. Introduction

Small family wineries play an important role in preserving the ethno-cultural landscape and constitute a significant element of the agricultural economy of the Republic of Moldova [42; 18]. Their importance is confirmed by official statistics: according to the National Bureau of Statistics [42], in 2024, small agricultural producers accounted for 63.4% of the gross grape harvest and 77% of the country's vineyard area. Over the past decade, a new wave of family winemaking entrepreneurship focused on producing high-quality wines has emerged in the country [44]. However, the development of this segment faces a persistent contradiction: while declaring a commitment to quality and traditions, these enterprises, according to empirical data, chronically underinvest in technological renewal and lack the managerial competencies necessary for effective market capitalization [32; 31; 17].

Contemporary research emphasizes that for small and medium-sized enterprises (SMEs), what becomes critically important is not merely possessing knowledge, but systematically managing it – its identification, preservation, transfer, and application [16; 12]. For family businesses, this task is complicated by the need to transfer informal, "tacit" knowledge between generations [13; 15]. Without formalized processes, accumulated experience risks being lost upon generational succession or the departure of a key employee [1].

Empirical data collected through an in-depth interview with the founders of the small family winery KaraGani Winery (September 2025) and a subsequent survey of 14 small farms

revealed a paradox: 100% of respondents cited lack of financial resources as the main obstacle to development. However, the successful case of KaraGani Winery demonstrates that access to grants alone is insufficient. A key role was played by the family's ability to institutionalize its knowledge – transforming oral rules into reproducible technological protocols, formalizing quality control procedures, and, ultimately, "packaging" unique experience into a brand eligible to use the protected geographical indication, recognized internationally through the IGP system[5; 14; 45]. This example, however, should not be considered typical; it represents more of an exception, demonstrating the development potential achievable through the synchronization of key factors.

The question arises: how can a small family winery synchronize investments in technology with the development of its ability to manage accumulated knowledge in order to achieve sustainable economic results? Approaches that reduce development to the sequence "investment → technology → quality → profit" [31] do not account for the specificity of small family businesses, where processes of knowledge management and intergenerational knowledge transfer play a critical role [16; 13]. As Meyer notes, any performance indicators degrade over time, and this problem is particularly acute in complex systems where non-financial indicators are difficult to aggregate [37].

The aim of this study is to construct and empirically substantiate a conceptual model for the modernization of a small family winery. This model views sustainable development as the result of the synchronous reinforcement of three interdependent basic elements – "nodes": technological, cognitive, and capital. Nodes are understood as key factors of production which, like points of support, must develop in a coordinated manner.

The scientific novelty of the research lies in the development of a conceptual modernization model which, unlike investment-oriented approaches, considers the sustainable development of a small family enterprise as the result of the synchronous reinforcement of three interconnected nodes. Novelty also lies in the interpretation of the system of protected geographical indications (IGP) as a dual institutional bridge, performing not only a market function but also a cognitive-disciplining one, stimulating the formalization of tacit family knowledge.

The key hypothesis is that a gap or asynchrony in the development of any node of the model blocks the enterprise's development, trapping it in one of several typical "traps" (for example, the "tradition trap" – possessing tacit knowledge but lacking modern technologies and the competencies to master them). The institutional mechanism IGP/DOP can function as a "bridge", converting modernization potential into market value [29; 14], while simultaneously acting as a disciplining factor, stimulating the formalization of internal processes [23]. The concept of strategic alignment, developed within enterprise architecture theory, confirms the necessity of synchronizing various subsystems to achieve sustainable development [36].

2. Theoretical Framework and Literature Review

This study draws upon three interrelated streams of literature: the economics of family business, theories of knowledge management in SMEs, and the theory of technological modernization. The synthesis of these streams provides the foundation for the conceptual modernization model proposed herein.

The primary asset of family wineries, which constitute the vast majority of producers in the global wine sector [4; 8], is the tacit knowledge accumulated over generations – an understanding of terroir, the nuances of technological processes, and the secrets of blending [1; 2; 23]. García et al. emphasize that heritage in family wineries serves as an active resource, shaping brand identity and competitive strategy [7]. However, this resource simultaneously

constitutes a source of vulnerability: research indicates that small family enterprises frequently encounter challenges related to knowledge continuity [12]. Tacit knowledge, if not formalized, risks being lost during generational succession [13; 15]. Consequently, a need arises to develop a set of processes enabling the family not only to preserve but also to augment its experience. In our model, this set of processes is termed the ***Cognitive Node***. To analyze this node, we employ the process model of knowledge management [12; 13; 14], which encompasses knowledge identification through experimentation and external contacts, and its codification through documentation – exemplified by the Antinori family's experience of transmitting winemaking expertise across 700 years to subsequent generations [23].

A critical element in the codification and application of knowledge is the system of management accounting and control. This can be viewed as the formalization of the financial aspects of tacit knowledge – understanding cost structures, the profitability of sales channels, and investment efficiency. As Breslavtseva and Sverchkova emphasize, for small enterprises, management accounting is not merely an option but an essential survival tool, enabling the assessment of the true financial position, the identification of underutilized resources, and informed decision-making [32].

For "tacit knowledge" to be transformed into a marketable product, two other components of the model are necessary: the material base and the financial resources for its creation. This leads us to conceptualize the enterprise as a system of three interrelated elements (nodes). Jeffrey Lamy, synthesizing practical experience in winemaking, demonstrates that creating a successful winery requires the simultaneous management of craftsmanship, marketing, and capital [17]. The ***Capital Node*** in our model represents the combination of financial resources and the managerial competencies required for their effective allocation. The ***Technological Node*** comprises the material base (equipment, tanks, bottling lines) that enables the formalization and standardization of production processes, thereby ensuring quality consistency [31]. The choice of technology is inextricably linked to the cognitive processes of goal-setting and planning, determining not only investment efficiency but also the requirements for personnel qualifications and knowledge formalization [30].

The synchronous development of the three nodes does not guarantee market success if the product's quality remains opaque to the consumer. An institutional mechanism is required to convert internal efforts into market value. One such mechanism is provided by geographical indication (GI) system, which reduce information asymmetry and generate credible quality signals [24, 25, 42, 43]. Research shows that the presence of a GI can increase producers' incomes and their willingness to invest in quality, acting not only as a signal to consumers but also as a coordination tool within the value chain [26; 27]. For small producers, GIs function as an "institutional elevator" (a term used by Menapace and Moschini [25]), enabling them to compete with larger players on the basis of authenticity and uniqueness. In the Republic of Moldova, the regulatory foundation for GIs is Technical Regulation No. 356/2015 [22]. However, the costs of compliance often remain a critical barrier for small producers [29].

In our model, IGP (Indicație Geografică Protejată) acts as a dual "institutional bridge." On the one hand, it performs an external, market function: it sets standards and requirements for the Technological Node and converts certified quality into market value (a price premiu m). On the other hand, it performs an internal, cognitive function: the requirements of IGP stimulate the formalization of knowledge – compelling the producer to document processes, maintain records, and standardize operations. It is important to emphasize that IGP fulfills this function only when the producer possesses sufficient resources and motivation to undergo certification. However, where certification has been achieved, it transforms tacit family knowledge into an explicit, reproducible, and controllable asset [23].

Thus, the synthesis of theories of family business, knowledge management, and institutional economics leads us to construct a conceptual model for the sustainability of small family wineries, in which development is possible only through the synchronous reinforcement of the three nodes – technological, cognitive, and capital – unified by the institutional bridge of IGP.

3. Research Methodology

The research design constitutes a qualitative study with elements of quantitative analysis. This approach allows for combining the depth of analysis of a single successful case with the possibility of preliminary generalization through a survey of a limited sample [38].

Qualitative component. An in-depth case study [40] was conducted based on the small family winery KaraGani Winery (Vulcănești, ATU Gagauzia). The primary data collection method was a semi-structured interview with both founders of the enterprise (duration: 2 hours 40 minutes), conducted on September 9, 2025. The interview aimed to reconstruct the history of the farm's development, identify key investment decisions, sources of financing, changes in technological processes, and knowledge management practices. An additional source of information was the analysis of internal documentation provided by the respondents: financial statements for 2023–2024, grant applications and reports, and conformity certificates, including the right to use the protected geographical indication IGP "Valul lui Traian" (Oficiul Național al Viei și Vinului, nr. 36AV of August 21, 2019) [45].

Quantitative component. In January–February 2026, a survey of small family winemaking farms was conducted. Respondents were selected using the snowball sampling method: initial contacts were established through industry associations and the personal networks of the founders of KaraGani, who subsequently recommended other producers. All participants were informed of the research objectives and provided informed consent for the processing and publication of anonymized data. The inclusion criterion was operation as a family enterprise with a production volume characteristic of small farms (predominantly up to 10,000 liters per year, confirmed by 93% of respondents). This approach yielded a sample of 14 farms, located primarily in the southern region of the Republic of Moldova, which introduces a risk of systematic bias. This limitation is taken into account when interpreting the results.

The questionnaire included closed-ended questions concerning production structure, experience with grants, and certification, as well as open-ended questions about perceived barriers to development. The sample is not representative of the entire population of small wineries in Moldova; however, it enables the identification of persistent patterns and the formulation of hypotheses for further research.

To verify the obtained data and broaden the research context, an official request was submitted to the National Office of Vine and Wine (NOVW) for statistics on producers certified to produce IGP wines. The response received (February 2026) contains a complete list of such producers with their years of registration, allowing for a comparison of the survey results with industry-wide dynamics.

Analytical methods. Interview processing was carried out using thematic analysis [41]: recordings were transcribed, coded, and key themes corresponding to the model's nodes were identified. Quantitative survey data were processed using descriptive statistics (calculation of proportions and distributions). Open-ended survey questions were analyzed thematically: responses were grouped into meaningful categories (e.g., "certification cost," "procedural complexity"), and the frequency of mentions for each category was calculated.

Cross-analysis was conducted by comparing respondents' answers with the patterns identified in the case study. For each farm, an assessment was made of whether it fell into situations analogous to the "traps" identified in the model:

- lack of financing despite possessing knowledge (well-developed cognitive node, underdeveloped capital node);
- receipt of grants despite weak accounting and marketing practices (well-developed capital node, lagging cognitive node);
- non-utilization of IGP despite producing a quality product (existing potential, absence of institutional capitalization).

This comparison provided preliminary confirmation of the conceptual propositions of the model on empirical material, acknowledging the limitations of the sample size.

4. Research Results: The KaraGani Winery Case Study

The evolution of KaraGani Winery (vineyard area: 7 hectares) allows us to trace how each key activity progressively developed particular nodes of the model. Until 2013, the enterprise relied exclusively on manual labor and empirical control of vinification, which ensured the accumulation of tacit family knowledge.

A turning point came in 2013 with the receipt of a grant from the Government of Japan (JNPGA) amounting to €60,000 (co-financing conditions: 80/20). This enabled the acquisition of a set of technological equipment (destemmer-crusher, membrane press), ensuring fraction separation and temperature control - a transition to controlled technological processes (the technological node). Between 2017 and 2021, a series of grants (USAID, SARD, the Executive Committee of Gagauzia, SLPA, ODA), totaling over 1.8 million lei, funded the acquisition of aging equipment (barriques), solar panels, infrastructure elements for enotourism, and backup power equipment.

An interview with the founders, conducted on September 9, 2025, reveals that modernization did not reject traditional knowledge but rather formalized and enhanced it. This process is most vividly illustrated by three technological decisions.

Example 1: Fraction separation. The family's traditional empirical rule stated: "never mix free-run juice with press fractions." The acquisition of a five-stage pneumatic press allowed this rule to be implemented technically: the press is set to select only fractions at a pressure not exceeding 0.2 bar. Technology did not abolish tradition but translated it from the realm of "sensation" into the realm of a reproducible standard.

Example 2: Oxidation control. Family practice included an empirical taboo against introducing baker's yeast into the must, explained by the need to avoid undesirable fermentation. During modernization, this rule was reinterpreted and replaced by creating an inert CO_2 environment at all stages of processing—from crushing to aging. The traditional prescription received scientific justification and technical implementation.

Example 3: Malolactic fermentation. Entering international markets required bringing wines up to premium standards. The introduction of bacterial cultures for malolactic fermentation made it possible to reduce the high titratable acidity of red wines from 8 to 5.5 units, significantly soften the tannic structure, and produce a product competitive in European competitions.

In 2019, KaraGani Winery, as one of the first small enterprises in the IGP region "Valul lui Traian" (Moldova), obtained the right to use the geographical indication for varietal wines, particularly the high-quality local variety Feteasсă Neagră [45]. According to the founders, this decision was strategic, driven by the aspiration to move beyond the tourist flow, work with restaurants, and participate in international exhibitions on an equal footing with European producers. The IGP status was perceived not as a bureaucratic burden, but as recognition that

the wine embodies the uniqueness of place and territory. The histories of renowned wine houses, such as Antinori, confirm that the ability to formalize and transmit knowledge across generations underpins long-term success [23].

The analysis of the winery's cognitive evolution, reconstructed from interview materials, allows us to detail the transition from informal family practices to documented procedures. The main drivers of this transition were: participation in NOVW seminars and contacts with equipment manufacturers; the transformation of empirical rules into standard protocols and the implementation of regular laboratory controls; and the involvement of children in production processes and the consulting of colleagues. The culmination of this process was the creation of the premium wine line "CIOTRA" and successful IGP certification.

The dynamics of financial indicators for 2023–2024, presented in Table 1, demonstrate the business's increased resilience to external shocks (in this case, the 2024 crop failure). Net profit decreased by 45% (from 86,368 to 47,395 lei); however, return on sales declined only slightly - from 7.83% to 6.01%. The relatively low level of profitability is explained by the active investment phase of development (equipment acquisition, certification), characteristic of small enterprises at the scaling stage.

The key factor allowing the volume loss to be offset was the increase in the average sales price from 52.9 lei/l in 2023 to 82.1 lei/l in 2024, i.e., by 55.2%. To assess the extent to which this increase was due to reputational factors (IGP status, international awards) rather than general market conditions, we consulted data from the National Bureau of Statistics [42]. According to Table 5.8 of the Statistical Yearbook, the producer price index for grapes in 2024 was 137.2% relative to 2023, corresponding to an average industry-wide price increase of 37.2% [42].

Table 1. Key performance indicators of KaraGani winery (2023-2024)

KEY PERFORMANCE INDICATORS			
INDICATOR	***2023***	***2024***	***CHANGE***
PRODUCTION VOLUME, LITERS	**20 850**	**9 600**	**▼ 54%**
REVENUE, LEI	**1 102 742**	**788 253**	**▼ 29%**
NET PROFIT, LEI	**86 368**	**47 395**	**▼ 45%**
AVERAGE PRICE, LEI/L	**52,9**	**82,1**	**▲ 55,2%**
RETURN ON SALES, %	**7,83%**	**6,01%**	**▼ 1,82 p.p.**

Source: Internal financial records of KaraGani Winery

The ability to market a substantially smaller harvest at a significantly higher price enabled the enterprise not only to avoid losses but also to maintain profitability at a level close to that of the pre-crisis period.

5. Survey Results of Small Wineries

To verify the patterns identified in the case study, a survey of 14 small family farms in the southern region of Moldova was conducted in January–February 2026. Analysis of the obtained data allows us to characterize the state of the three nodes of the model within the studied sample and identify systemic barriers to development.

General characteristics of the sample. Among the respondents, farms with extensive winemaking experience predominate: 65% have been operating for over 15 years, and a further

14% for 10 to 15 years. At the same time, 93% produce up to 10,000 liters of wine per year, corresponding to the status of micro-enterprises. The assortment of the majority (79%) is limited to 2–3 wine varieties, and only 14% offer 5–10 items, including wines from indigenous grape varieties.

Cognitive node. The vast majority of respondents (86%) recognize quality as the main success factor, directly influencing sales, tourist attractiveness, and awards. However, this declared commitment to quality is not supported by institutional formalization: 64% of farms do not certify each batch of production, relying on traditional methods and customer feedback. Only 21% undergo mandatory certification for each batch, and a further 14% certify only new products when the recipe changes.

Technological node. This node is characterized by significant backwardness: 71% of respondents work "the old-fashioned way," without implementing new technologies over the past 10 years. Only 29% have acquired modern equipment (presses or temperature-controlled tanks). Moreover, 57% are ready to modernize only on the condition of receiving a grant, which creates a risk of a "capitalization gap." A further 29% do not yet plan changes, considering traditional methods sufficient, and only 7% are oriented towards digitalization (vineyard sensors, online sales).

Capital node. This node demonstrates a high dependence on external financing with limited access to it. Experience in receiving grants is held by 29% of farms; 71% have never received either grants or loans, financing development exclusively from their own funds. None of the respondents have used credit resources (100%). At the same time, all respondents (100%) cited lack of finance as the main barrier to implementing new technologies.

Analysis of the open-ended questions revealed the main obstacles to obtaining IGP status, which explain why 64% of farms do not certify their production despite recognizing the importance of quality. The high cost of certification and analyses was noted by 73% of respondents, and the complexity and bureaucratization of procedures by 64%. More than half (55%) indicated a lack of information about the certification process, and 45% are unsure of the economic return on investment.

To verify the obtained conclusions at the industry level, an official request was sent to the National Office of Vine and Wine. The data received (February 2026) contain a complete list of producers certified to produce wines with a protected geographical indication (IGP), including their years of registration. Figure 1 presents the dynamics of this process.

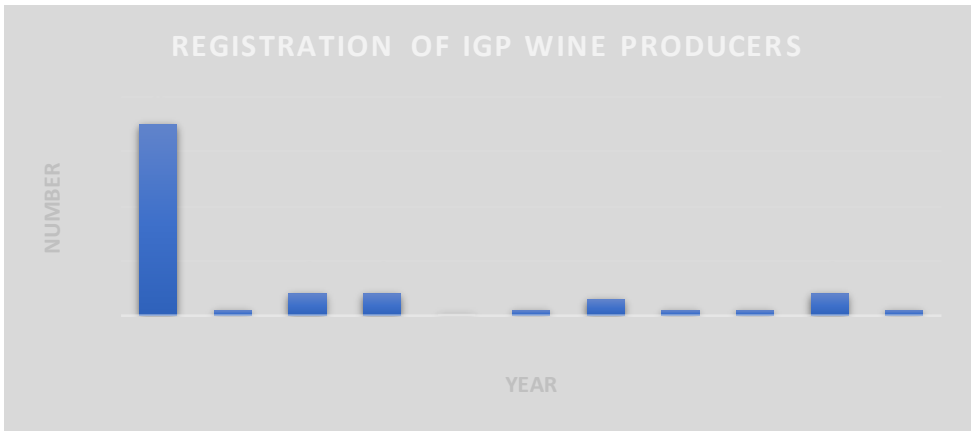


Fig. 1. Dynamics of IGP wine producer registration in the Republic of Moldova

The graph clearly shows two stages in the institutional development of the IGP system in Moldova. The first stage (2016) is characterized by a mass entry of producers - in a single year, 35 enterprises received IGP status, accounting for 66.7% of all those certified by 2026. The second stage (2017–2026) is marked by isolated additions: on average, 1-4 new producers per year, with no growth at all in 2020.

6. Conceptual Model of Modernization

The analysis of the KaraGani Winery case study and the survey data from small wineries allows us to formalize a model for the sustainable development of a small family winery – the "MODERNIZATION TRIAD" (Fig. 2). The model proceeds from a simple yet often overlooked thesis: the sustainability of an enterprise depends not on a single resource, but on the synchrony of all, in this case, three interconnected nodes:

• **TECHNOLOGICAL NODE** – the material and technical base (equipment, process control, quality consistency);

• **COGNITIVE NODE** – the knowledge management system (identification and codification of tacit knowledge, experience sharing, continuity);

• **CAPITAL NODE** – financial resources and managerial competencies (access to grants, management accounting, profitability).

Nodes are understood as key factors of production and development; the lagging of any one of them renders the system unstable. The institutional mechanism for synchronization is the system of protected geographical indications (IGP/DOP), performing the function of a dual bridge. This structure resonates with the ideas of strategic alignment developed within enterprise architecture theory, where the coordinated development of various subsystems is considered a necessary condition for effectiveness [36].

The structure of the model is represented by a diagram of elements reflecting various aspects of the functioning of a family winemaking enterprise, presented by three basic nodes and the IGP institutional bridge block. Numbers 1–6 on the bidirectional arrows between the nodes indicate inter-nodal influences:

- ① — knowledge influences the choice of technologies and equipment;
- ② — technology enables the formalization and standardization of tacit knowledge;
- ③ — financial resources enable technological renewal;
- ④ — effective technologies increase profitability;
- ⑤ — investments are directed towards staff training and development;
- ⑥ — managerial competencies enhance the efficiency of capital utilization.

Bidirectional dashed arrows between each node and the "IGP Institutional Bridge" block reflect mutual influence: the level of node development determines the possibility of obtaining certification (technologies must meet standards, knowledge must be formalized, capital must allow covering the procedures); in turn, IGP influences the nodes by setting standards (technological node), stimulating knowledge formalization (cognitive node), and converting certified quality into market value (capital node).

The broad arrow leads to the ultimate result of the synchronous development of all three nodes through the **IGP INSTITUTIONAL BRIDGE** - sustainable development, understood as the enterprise's ability to maintain profitability and reputation under conditions of external shocks (crop failure, market volatility).

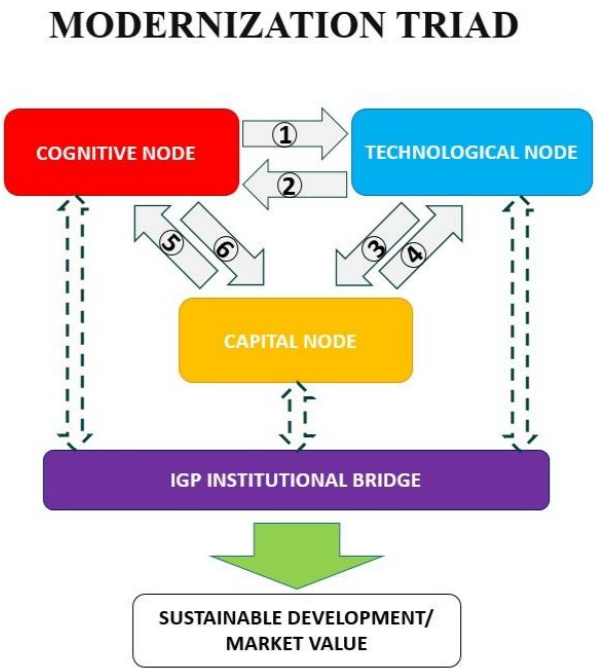


Fig. 2. Conceptual model for the modernization of a small winery

The institutional mechanism capable of synchronizing the development of all three nodes and "extracting" the enterprise from the described traps is the system of protected geographical indications (IGP/DOP). Within the framework of the model, IGP is interpreted as **a dual institutional bridge**:

- ***the external bridge*** – *sets standards and requirements for the technological node and converts certified quality into market value, providing a price premium.*
- ***the internal bridge*** – *certification requirements compel the producer to formalize knowledge: documenting processes, standardizing operations, thereby creating a foundation for intergenerational knowledge transfer.*

It is important to emphasize that IGP performs this function only under the condition that access barriers are overcome: where certification has been achieved, it triggers the mechanism of node synchronization, transforming family experience into a market asset.

While the KaraGani Winery case demonstrates that the synchronous development of the three nodes combined with institutional capitalization through IGP ensures sustainability even in unfavorable years, the survey data from other wineries record the opposite situation: asynchrony or a gap in the interaction of the nodes keeps enterprises in one of three typical "traps," each of which has empirical confirmation (Table 2).

Thus, the "MODERNIZATION TRIAD" describes a model for the sustainable development of a small family winery as a dynamic process of synchronous reinforcement of the three nodes—technological, cognitive, and capital—united by the institutional bridge of IGP, which standardizes production, formalizes knowledge, and capitalizes quality into market success.

Table 2. Typology of traps (GAPS) in the modernization triad

TYPOLOGY OF MODERNIZATION TRAPS		
TRAP/GAP	**CHARACTERISTICS**	**EMPIRICAL BASIS**
"TRADITION TRAP"	Unstable quality – inability to scale	**71%** of the sample work "the old-fashioned way"
"CAPITALIZATION GAP"	Risks of relying solely on external financing	**57%** of respondents are ready to modernize only if they receive a grant
"IGP BARRIER"	Lack of institutional recognition blocks access to price premium	**64%** do not certify their products

Source: Survey of small wineries (n=14)

7. Discussion of Results and Conclusyiona

This study contributes to the discourse on pathways for modernizing small family agricultural enterprises. The proposed "triad" model advances resource-based approach ideas within the wine sector context. While previous research often focused on individual factors — technological renewal [31; 13], knowledge management [16; 12], or financial mechanisms [32; 17]—our model demonstrates their systemic interrelationship, showing that each of these factors yields effects only in conjunction with the other two.

The key theoretical conclusion is that the sustainability of a small family enterprise is ensured not by the maximal development of a single resource, but by the synchronization of the three nodes. Empirical evidence confirms: the presence of grants (capital) without formalized knowledge (cognitive node) leads to a "capitalization gap"; rich traditions without technologies reinforce the "tradition trap." Thus, we refine Lamy's thesis on the necessity of simultaneously managing craftsmanship, marketing, and capital [17], and also develop the ideas of De Massis et al. on the specificity of innovation in family firms [5], concretizing the synchronization mechanisms for small winemaking enterprises. Our approach also resonates with the concept of strategic alignment developed within enterprise architecture theory, where the coordinated development of various subsystems is considered a necessary condition for effectiveness [36].

The second significant finding concerns the dual role of IGP. In the literature, geographical indications have traditionally been viewed as a marketing tool and a means of protection against counterfeiting [24; 25; 42]. Our research shows that for family wineries, the process of preparing for IGP/DOP certification performs an equally important internal cognitive function, compelling the formalization of tacit knowledge and creating a foundation for intergenerational knowledge transfer. IGP acts not merely as a certificate, but as a managerial disciplining mechanism, opening a new line of inquiry: the study of geographical indication systems as knowledge management tools in small businesses. The historical experience of leading wine houses, such as Antinori, confirms that the ability to formalize and transmit knowledge across generations underpins long-term success [23].

References

1. Polanyi M. Personal Knowledge: Towards a Post-Critical Philosophy. London: Routledge & Kegan Paul, 1958. 428 p.

2. Nonaka I., Takeuchi H. The Knowledge-Creating Company: How Japanese Companies Create the Dynamics of Innovation. New York: Oxford University Press, 1995. 304 p.
3. Nonaka I., von Krogh G. Tacit Knowledge and Knowledge Conversion: Controversy and Advancement in Organizational Knowledge Creation Theory // Organization Science. 2009. Vol. 20. No. 3. P. 635-652.
4. De Massis A., Frattini F., Lichtenthaler U. Research on Technological Innovation in Family Firms: Present Debates and Future Directions // Family Business Review. 2013. Vol. 26. No. 1. P. 10-31.
5. De Massis A., Frattini F., Majocchi A., Piscitello L. Family firms in the global economy: Toward a deeper understanding of internationalization determinants and mechanisms // Global Strategy Journal. 2018. Vol. 8. No. 1. P. 3-21.
6. Chrisman J.J., Chua J.H., De Massis A., Frattini F., Wright M. The management of family firms: A review and research agenda // Journal of Family Business Strategy. 2015. Vol. 6. No. 4. P. 211-226.
7. García J., López E., Fernández R. Heritage as a source of competitive advantage in family wineries: A longitudinal study // Journal of Family Business Strategy. 2022. Vol. 13. No. 4. P. 1480-1495.
8. García F.A., et al. From heritage to sustainability: a systematic review of family business in wine industry // Journal of Family Business Management. 2023. Vol. 15. No. 6. P. 1478-1502.
9. Osipov E.M., Dudina O.M., Sazonova A.P. Family entrepreneurship in modern Russia: problems and prospects. Moscow: Prometheus, 2024. 320 p. (In Russian)
10. Gergert D.V., Ponomareva T.V. Family business as an object of interdisciplinary research: a review of methodological approaches // Perm University Herald. Economics. 2018. Vol. 13. No. 1. P. 90-105. (In Russian)
11. Stamov I. Comparative Analysis of Legislative Support for the Organization of a Family Business // Proceedings of the International Scientific Conference "Economic Welfare through Knowledge Sharing". UNWE, 2024. P. 1-8.
12. Durst S., Edvardsson I.R. Knowledge management in SMEs: a literature review // Journal of Knowledge Management. 2012. Vol. 16. No. 6. P. 879-903.
13. Scuotto V., Del Giudice M., Bresciani S., Meissner D. Knowledge management in SMEs: a state-of-the-art review // Journal of Knowledge Management. 2021. Vol. 25. No. 11. P. 1-23.
14. Sandkuhl K., Smirnov A. Knowledge management in small and medium-sized enterprises: a systematic review // Journal of Knowledge Management. 2023. Vol. 27. No. 6. P. 1618-1646.
15. Bratianu C., Bejinaru R. The theory of knowledge fields: thermodynamics approach // Proceedings of the 14th European Conference on Knowledge Management. 2013. P. 90-97.
16. Sima V., Gheorghe I.G., Subić J., Nancu D. Influences of the Industry 4.0 Revolution on the Human Capital Development and Consumer Behavior: A Systematic Review // Sustainability. 2020. Vol. 12. No. 10. P. 4035.
17. Lamy J. The Business of Wine: An Encyclopedia. Santa Barbara: Greenwood Press, 2020. 320 p.
18. Santini C., Cavicchi A., Casini L. Sustainability in the wine industry: key questions and research trends // Agricultural and Food Economics. 2013. Vol. 1. No. 1. P. 1-14. URL: https://link.springer.com/article/10.1186/2193-7532-1-9 (accessed: 20.02.2026)
19. Broccardo L., Culasso F., Truant E. Sustainability in the wine sector: The role of innovation and dynamic capabilities // Sustainability. 2019. Vol. 11. No. 15. P. 4036.
20. Režek J., et al. Strategic Resilience in Small Wineries: A European Perspective // Sustainability. 2024. Vol. 16. No. 7. P. 2857.
21. Nave E., et al. Zooming-in small family wineries: exploring service innovation // International Journal of Wine Business Research. 2023. Vol. 36. No. 4. P. 613-631.
22. De Bernardi P., Bertello A. Holding back the years: Tradition and innovation in family wineries // ResearchGate. 2022. URL: https://www.researchgate.net/publication/361583885_Holding_back_the_years_Tradition_and_innovation_in_family_wineries (accessed: 20.02.2026)
23. Antinori P. The Hills of Chianti: The Story of a Tuscan Winemaking Family, in Seven Bottles. New York: Rizzoli Ex Libris, 2018. 176 p.
24. Josling T. The War on Terroir: Geographical Indications as a Transatlantic Trade Conflict // Journal of Agricultural Economics. 2006. Vol. 57. No. 3. P. 337-363.
25. Menapace L., Moschini G. Quality certification by geographical indications, trademarks and firm reputation // European Review of Agricultural Economics. 2012. Vol. 39. No. 4. P. 539-566.

26. Antonioli F., Ciaian P., Baldoni E. The impact of geographical indications on farms' performance. An empirical analysis of the EU vineyard sector. JRC Technical Report. Luxembourg: Publications Office of the European Union, 2024. 45 p.
27. Poetschki K., Peerlings J., Dries L. The impact of geographical indications on farm incomes in the EU olives and wine sector // British Food Journal. 2021. Vol. 123. No. 13. P. 278-298.
28. Costanigro M., et al. The Information Value of Geographical Indications: An Experimental Auction with U.S. Consumers // Food Policy. 2024. Vol. 122. P. 102567.
29. Simion T. Protecția indicațiilor geografice în Republica Moldova: provocări și perspective pentru micii producători // Intellectus. 2023. Nr. 1. P. 42-48.
30. Sukharev O.S. Theory of technological modernization: institutional and structural analysis // Economic Science of Modern Russia. 2022. No. 4 (99). P. 7-26. (In Russian)
31. Goncharov V.N., Ivanova E.V. Technological modernization of agricultural enterprises: factors and efficiency // Economics of Agriculture in Russia. 2021. No. 3. P. 22-27. (In Russian)
32. Breslavtseva N.A., Sverchkova O.F. Management accounting in small enterprises: problems and solutions // Modern Economics: Problems and Solutions. 2017. No. 6 (90). P. 75-82. (In Russian)
33. Zemtsov S.P., Chepurenko A.Yu., Barinova V.A. Small and medium-sized entrepreneurship in Russia: institutional barriers and the role of regional factors // Voprosy Ekonomiki. 2021. No. 3. P. 25-47. (In Russian)
34. Van de Ven A.H., Poole M.S. Explaining development and change in organizations // Academy of Management Review. 1995. Vol. 20. No. 3. P. 510-540.
35. Hax A., Wilde D. The Delta Model: Adaptive Management for a Changing World // MIT Sloan Management Review. 1999. Vol. 40. No. 2. P. 11-28.
36. Maidanova S.A. Modeling the IT architecture of a global company based on a balanced scorecard: PhD thesis. St. Petersburg: SPbPU, 2020. 177 p. (In Russian)
37. Meyer M.W. Rethinking Performance Measurement: Beyond the Balanced Scorecard. Cambridge: Cambridge University Press, 2002. 202 p. (English original; Russian translation: Moscow: Vershina, 2004)
38. Creswell J.W., Poth C.N. Qualitative Inquiry and Research Design: Choosing Among Five Approaches. 4th ed. SAGE Publications, 2018. 488 p.
39. Creswell J.W., Plano Clark V.L. Designing and conducting mixed methods research. 3rd ed. SAGE Publications, 2017. 520 p.
40. Stake R.E. The Art of Case Study Research. SAGE Publications, 1995. 192 p.
41. Braun V., Clarke V. Using thematic analysis in psychology // Qualitative Research in Psychology. 2006. Vol. 3. No. 2. P. 77-101.
42. Statistical Yearbook of the Republic of Moldova 2025. National Bureau of Statistics. Chișinău, 2026. 560 p.
43. Agricultural activity of small producers in the Republic of Moldova, 2025 edition. National Bureau of Statistics. Chișinău, 2026. URL: https://statistica.gov.md/ (accessed: 20.02.2026)
44. Organization for Entrepreneurship Development. Progress report on the implementation of state programs and assistance projects: First half of 2025. Chișinău: ODA, 2025. 67 p. URL: https://www.oda.md/files/rapoarte/2025/RAPORT%20privind%20evaluarea%20progresului%20pentru%20sem.%20I,%20anul%202025.pdf (accessed: 20.02.2026)
45. State Agency for Intellectual Property (AGEPI), official page. URL: https://agepi.gov.md/sites/default/files/bopi/BOPI_12_2022.pdf (accessed: 20.02.2026)

Information and Telecommunication Technologies as the Basis of Modern Cyber Security

Liudmyla Rybalchenko

University of Customs and Finance, Dnipro, Ukraine
luda_r@ukr.net

Abstract. The article analyzes the role and place of information and telecommunication technologies (ITT) in the structure of modern cybersecurity systems. The evolution of architectural approaches is studied: from classical perimeter defense to the Zero Trust model. Key components of ITT, such as cloud computing, artificial intelligence, cryptographic protocols, and zero trust network architectures, are analysed. Particular attention is paid to the transformation of protection methods under the influence of the growing complexity of cyber threats. An analysis of the integration of artificial intelligence and machine learning as an intellectual core for analysing large arrays of security events in real time is performed. The use of blockchain technology to ensure the immutability of audit logs is considered. Based on the analysis of major cyberattacks during 2022–2025, critical technical errors are identified, and ways to strengthen cyber resilience through the implementation of EDR/XDR systems and response automation are proposed.

Keywords: cyber defense, information and telecommunication technologies, artificial intelligence, machine learning, Zero Trust, blockchain, information security, monitoring.

1. Introduction

Today, the world is undergoing a phase of total digitalisation, characterised by the penetration of information and telecommunication technologies (ITT) into all areas of life: from public services and medicine to critical infrastructure management. The concepts of the 'Smart City' and the digital economy are based on the continuous exchange of data. However, this global interdependence creates unprecedented challenges for national and international security. Cyber defence is no longer an optional extra but has become an integral part of the architecture of any information system.

Global digitalisation is an irreversible process that has brought enormous benefits to humanity, but at the same time has made it vulnerable. New challenges require a transition from 'patching vulnerabilities' to building resilient systems where protection is integrated into the IT architecture itself at the design level (Security by Design).

Global digitalisation has led to exponential growth in the number of connected devices. The introduction of Internet of Things (IoT) and 5G technologies means that the entry point for a hacker can now be not only a company's server, but also a 'smart' temperature sensor, surveillance camera or medical implant. Each new node in the network expands the overall 'attack surface' that needs to be protected.

The modern challenges lie in the fact that cyber attacks are no longer limited to virtual space. The digitisation of critical infrastructure (power grids, water supply systems, transport) means that a successful intrusion into an IT system can lead to real man-made disasters. Cyber weapons have become a tool of geopolitical influence, as clearly demonstrated by the events of 2022–2025.

2. Research objectives and methods

The aim of the article is a comprehensive study of the role of information and telecommunication technologies in forming resilient cyber defense systems, as well as an analysis of the effectiveness of integrating artificial intelligence, Zero Trust architecture, and blockchain technology to counter modern cyber threats in the context of global digitalization.

To achieve the set goal, a set of scientific methods was used to ensure objectivity and depth of analysis:

1). System analysis method: used to study ITT as an integral structure, where each component (network, endpoint, cloud) plays a specific role in the overall security system.

2). Comparative-typological method: used to compare traditional perimeter protection models with modern Zero Trust architectures, as well as to analyse different data formats (IDMEF, JSON, CEF).

3). Statistical and descriptive methods: used to analyse large-scale cyberattacks in the period 2022–2025 to identify patterns of technical errors and their impact on the scale of incidents.

4). Process modelling: used to describe the mechanisms of the intelligent core based on machine learning algorithms (e.g., Random Forest) in the process of detecting anomalies in network traffic.

5). Deduction and generalisation method: allowed us to formulate a comprehensive concept for the promising development of autonomous cyber defence systems based on the studied technological trends (blockchain, AI, 5G).

3. Presentation of the main research material

Traditional defence methods based on human analysis face the challenge of speed. Malicious actors are increasingly using artificial intelligence to automatically search for vulnerabilities in software code, generate personalised phishing (Deepfakes) and create viruses that change their structure (polymorphic code) during propagation.

There are issues of trust and digital transparency, as in a globalised world, data is often stored on servers in different countries (cloud computing). This creates a challenge for legal and technical control to ensure data security. This stimulates the development of confidential computing technologies and the localisation of critical network segments.

Given the rapid digitalisation of society, the human factor plays an important role in the digital environment, which is the weakest link. Digitalisation is outpacing the level of cyber hygiene among the population. It makes services as accessible as possible: one-click payment, biometrics, cloud storage.

The problem is that users get used to comfort, perceiving digital services as 'magic' that works by itself. This lulls them into a false sense of security. For example, a person may use one easy password for all services in order to avoid unnecessary inconvenience [7].

Previously, phishing emails were easy to spot due to errors in the text. Today, attackers use neural networks. Hackers imitate the voice of a boss or the video image of a relative to trick a person into transferring money or providing access to the network. Artificial intelligence analyses the victim's social networks and generates messages that a person is 90% likely to believe. The cyber hygiene of the population simply cannot keep up with such a rapid pace of manipulation.

Company employees often use personal messengers or cloud services for work documents because it is 'faster.' They do not realise that they are removing corporate data from the protection of SIEM or DLP systems. Here, the digitalisation of processes (the transition to the cloud) has outpaced the understanding of the risks of data leakage.

According to international reports (including IBM and Verizon), between 82% and 90% of all successful cyber attacks start with the 'human factor' [6]. The average time it takes to train a user in new security rules is 6-12 months, while new types of attacks appear every week.

Since it is impossible to teach 100% of the population perfect cyber hygiene, modern cyber protection shifts the responsibility to technologies, which include:

- Passwordless (passwordless login), i.e. replacing passwords with hardware keys (FIDO2) or biometrics so that the user cannot 'accidentally' give their password to a phisher;
- Zero Trust - the system does not trust the user even after login, constantly checking their actions for anomalies;
- automated sandboxes - checking all attachments in email with AI algorithms before the user has time to open them.

Thus, poor cyber hygiene is a link in the chain of modern cyber threats. Therefore, the architecture of modern information and telecommunications technologies must be based on the principle of 'resilience to human error,' where the protection system is capable of automatically neutralising the consequences of ill-considered user actions.

Modern cyber protection based on the CIA triad is important: confidentiality (protection against unauthorised access); integrity (protection against unauthorised data modification); availability (ensuring uninterrupted access to services). Information and telecommunications technologies ensure the technical implementation of this triad through the introduction of automated monitoring and access control systems.

An analysis of major cyberattacks in recent years shows that even the world's most powerful corporations are vulnerable due to basic technical miscalculations. Below is an analysis of the most high-profile incidents of the period 2022–2025.

Looking at the attack on MGM Resorts in 2023, which was one of the most high-profile attacks and paralysed hotels in Las Vegas for several weeks, it should be noted that the Scattered Spider group gained access to the internal network through a regular phone call to the IT help desk, posing as an employee [3]. The technical flaws in the attack were the lack of strict identity verification, i.e. the support service did not have a protocol for visual or cryptographic identity verification when resetting passwords. The account privileges were excessive. Once they gained access to a single account, the hackers were able to quickly elevate their privileges to Okta (Identity Provider) administrator level, giving them control over the entire infrastructure.

Another example of an attack that took place in 2023 was the MOVEit hack (2023) - a zero-day vulnerability that affected thousands of organisations around the world, including US government agencies and major banks [3].

Hackers exploited a critical SQL injection vulnerability in the MOVEit Transfer file transfer software. Analysing the mistakes that were made, it was a lack of proper protection when logging into the system. The software did not properly filter input data in SQL queries, allowing hackers to execute arbitrary commands in the database. Direct access to the databases was gained, allowing attackers not only to read but also to download huge amounts of confidential information.

Another example was the attack on Snowflake in 2024 due to a large-scale leak of customer data from the Snowflake cloud platform (including Ticketmaster and Santander Bank) [4]. Hackers stole data from Snowflake customer accounts using previously stolen credentials.

Technical errors include the lack of mandatory multi-factor authentication (MFA), meaning that many customer accounts were protected only by a login and password. In today's environment, this is considered a critical configuration error.

In addition, the long lifetime of session tokens meant that sessions did not end automatically, giving attackers time to download terabytes of data without re-authentication.

In 2025, there was a surge in attacks where AI-Driven Phishing generated perfect phishing messages based on open source intelligence (OSINT) [5]. The mistake was placing too much trust in traditional email filters. Old email protection systems look for known malicious links, but miss AI-generated emails that do not contain viruses, but only persuade the user to voluntarily transfer data.

Thus, the scale of attacks is usually determined not by the genius of hackers, but by three factors: human - staff unpreparedness for social engineering, weak identity management (IAM) - lack of MFA and the principle of least privilege, and lack of visibility - companies learn about breaches too late due to the lack of configured monitoring and event correlation (SIEM).

4. Conclusions

Thus, reliable cyber protection requires a technological transformation that will apply automated threat detection and neutralisation in automated systems with information and telecommunications technologies for data transmission and processing.

With millions of security events generated every second, the introduction of artificial intelligence and machine learning is a critical necessity. Only algorithms are capable of detecting complex anomalies and zero-day attacks in real time that remain invisible to humans.

Traditional protection is no longer effective. The transition to Zero Trust architecture minimises the risks associated with account compromise and internal threats through constant verification of each network entity, which is also an advantage of this zero trust concept.

The use of decentralized registries for storing audit logs solves the problem of 'trust in the administrator.' Blockchain guarantees the immutability of the evidence base, which is a key factor for successful investigation of cybercrimes (Digital Forensics). Despite the high technological sophistication of modern ITT, configuration errors and social engineering remain the main vulnerabilities. Effective cyber protection is only possible when advanced technologies (EDR/XDR/SIEM) are combined with ongoing staff training and strict adherence to security policies.

References

1. On the fundamental principles of ensuring cybersecurity in Ukraine: Law of Ukraine dated 05.10.2017 p. № 2163-VIII, https://zakon.rada.gov.ua/laws/show/2163-19
2. Sarker I. H. Cybersecurity Data Science: Best Practices, Methods and AI-Based Solutions. Springer Nature, 2022. 245 p.
3. Zero Trust Architecture : NIST Special Publication 800-207 / S. Rose, O. Borchert, S. Mitchell, S. Connelly. Gaithersburg : National Institute of Standards and Technology, 2020. 59 p.
4. Cost of a Data Breach Report 2024. IBM Security, 2024. https://www.ibm.com/reports/data-breach
5. CrowdStrike. (2025). 2025 Global Threat Report: Confronting the Evolution of Cyber Adversaries. https://www.crowdstrike.com/global-threat-report/
6. IBM Security. (2025). Cost of a Data Breach Report 2025. https://www.ibm.com/reports/data-breach
7. Hrebenyuk A.M., Rybalchenko L.V. (2020). Fundamentals of information security management: training. manual. [Osnovy upravlinnya informatsiynoyu bezpekoyu: navch. posibnyk]. Dnipro: DDUVS, – 144 c.

Global Trends and National Realities in ESG Reporting

Irina Iavorschi, Liliana Lazari

Academy of Economic Studies of Moldova, Chisinau, Moldova
irinaiavorschi@gmail.com, Liliana.lazari@ase.md

Abstract. Environmental, Social, and Governance (ESG) reporting has emerged as a critical tool for organizations seeking to demonstrate their commitment to sustainable practices and social responsibility. As investor interest and regulatory pressures continue to rise, understanding the evolving trends and challenges in international ESG reporting has become essential. This research explores the global landscape of ESG reporting, examining the key trends, frameworks, and regulations shaping corporate disclosures. By analyzing the current state of ESG integration in the Republic of Moldova, this paper identifies the "implementation gap" between global mandates and national reporting capacities. Despite advances, challenges such as inconsistent data quality and the structural transition from fiscal-oriented to sustainability-oriented reporting persist. The study concludes that the successful integration of ESG reporting in Moldova is fundamentally contingent upon the transition from the fiscal-oriented National Accounting Standards (NAS) to the transparency-driven IFRS framework, as the latter provides the essential data architecture for credible sustainability disclosures.

Keywords: ESG Reporting, Sustainability, Global Standards, Reporting Frameworks

1. Introduction

Companies across industries now recognize the importance of aligning their operations with environmental, social, and governance (ESG) principles. Investing time, resources, and effort into embedding sustainability is a strategic requirement for accessing international capital markets; it's about redefining business models and seizing opportunities for growth and innovation. At its core, ESG reporting is the disclosure of environmental, social, and corporate governance data. This comprehensive approach allows companies to communicate their commitment to sustainability and responsible business practices to stakeholders, including investors, customers, and regulators. ESG reporting goes beyond financial metrics, providing insight into a company's broader impact on society and the environment. By disclosing ESG performance metrics, companies demonstrate their dedication to transparent and responsible business practices, fostering trust and credibility with stakeholders. Despite this global momentum, domestic investment in environmental protection remains significantly lower than European benchmarks. Both government and business spend too little on these purposes. In recent years, Moldova has allocated only 0.5% of GDP to environmental protection, while the EU average is 1.9%.[1, p.13] There are multiple reasons for under-investment, generally linked to financial constraints and the state of the Moldovan economy. Environmental protection is often sacrificed in favor of other political or economic priorities.

2. Methodology

This study adopts a qualitative, exploratory approach based on:

- Normative analysis: Reviewing EU Directives (CSRD/ESRS) and the Moldovan legislative framework (Environmental Strategy 2024-2030);

- Comparative benchmarking: Evaluating the reporting practices of national "Pioneering Entities" (MAIB S.A., Moldcell S.A., Purcari Wineries PLC) against ESRS standards;

- Diagnostic analysis: Investigating the implementation challenges of ESG reporting through a tiered approach, from regulatory inconsistencies to professional skill shortages;

- Gap analysis: Identifying the structural barriers in the transition from fiscal-oriented (NAS) to investor-oriented (IFRS/ESG) reporting.

3. Current trends in international ESG reporting

The year 2026 is positioned as a pivotal juncture in the European reporting timeline. Following the introduction of the "Omnibus I" package and the "Stop-the-Clock" Directive (Directive (EU) 2025/794), the EU has fundamentally recalibrated its implementation schedule to balance the urgent need for transparency with the necessity of maintaining corporate competitiveness [21]. The provisional agreement reached in December 2025 has significantly altered the size thresholds for mandatory reporting, focusing on entities with at least 1,000 employees and €450 million in annual net turnover [20, p.3].

Table 1. Recalibrated CSRD Implementation Timeline under the Omnibus I Package and Directive (EU) 2025/794

Reporting Category	New Size Thresholds (Omnibus I)	Initial Reporting Year	Report Publication Year
Wave 1 (NFRD Companies)	> 500 employees (historical)	2024	2025
Wave 2 (Large EU Undertakings)	> 1,000 employees & > €450M Turnover	2027	2028
Wave 3 (Non-EU Parent / Art 40a)	> €450M EU Group Turnover	2028	2029
Voluntary SMEs (VSME)	Below mandatory thresholds	Optional	Optional

Source: Prepared by the authors based on [20]

The "Stop-the-Clock" Directive effectively postpones existing reporting requirements for Wave 2 and Wave 3 entities by two years, providing a "transition year" in 2026 for these organizations to build the robust data-collection structures required for future compliance. Meanwhile, Wave 1 entities benefit from "Quick Fix" amendments adopted in July 2025, which allow them to maintain or even reduce the level of reporting applied in 2024 for their 2025 and 2026 reports, thereby reducing the immediate administrative burden during the phase-in period [20,p.1;5].

The evolution of ESG reporting has followed a gradual, tiered trajectory, with large companies leading the adoption process and small and medium-sized enterprises (SMEs) progressively following suit. Initially, ESG disclosure was largely voluntary and concentrated among large, publicly listed entities with strong exposure to international markets and investor scrutiny. These companies were the first to align with early frameworks such as the GRI, TCFD, or SASB, driven by reputational and regulatory incentives. The introduction of the EU Non-Financial Reporting Directive (NFRD), and more recently the Corporate Sustainability Reporting Directive (CSRD), marked a significant shift toward mandatory ESG disclosure. Under CSRD, large entities—including those already covered by NFRD—are required to

publish ESG reports aligned with the European Sustainability Reporting Standards (ESRS) starting from the 2024 financial year, with a phased extension to non-listed large companies and SMEs. Listed SMEs will become subject to mandatory reporting obligations from the 2026 financial year, with optional deferrals until 2028.

This phased implementation reflects both the growing importance of ESG transparency and the structural differences in reporting capacity across company sizes. While large entities generally possess the financial and human resources necessary to integrate ESG into their governance systems, SMEs face notable challenges—ranging from limited internal expertise and high compliance costs to difficulty accessing ESG-specific assurance services. Nevertheless, SMEs are increasingly impacted by indirect pressures via supply chains, investor expectations, and regulatory alignment, making ESG integration a strategic imperative across all company tiers.

Table 2. Timeline of ESG Reporting Expansion: Large Companies vs. SMEs (EU, 2014–2028)

Year / Phase	Large Entities	Small and Medium-sized Enterprises (SMEs)
2014–2017	Voluntary ESG reporting, primarily by large listed companies	Minimal or negligible ESG reporting practices
2018	NFRD becomes effective – mandatory for large public-interest entities	SMEs not covered by NFRD
2021	CSRD proposed – extends obligations to all large companies	Upcoming applicability announced for listed SMEs
2024 (FY)	Mandatory reporting under CSRD and ESRS (reports published in 2025)	Not yet mandatory, but indirect pressures increase
2026 (FY)	Majority of large companies report under ESRS	Listed SMEs begin mandatory reporting under CSRD (reports in 2027)
2028 (FY)	Expanded ESG audit requirements introduced ("reasonable assurance")	Final deadline for listed SMEs; potential voluntary extension

Source: Prepared by the authors based on [11,12,13,14]

Although the CSRD Directive mandates sustainability reporting for approximately 50,000 companies within the European Union, recent data indicate that only a portion of these entities currently comply with the requirements. ESG reporting is predominantly observed among large companies, with approximately 82–90% of entities within the N100 category already disclosing non-financial information. However, surveys reveal that less than 5% of the companies targeted in the first phase of CSRD implementation are fully prepared to meet the new standards. This discrepancy highlights a significant gap between the legal obligation and the actual capacity for compliance, underscoring the need for concerted internal adaptation efforts and institutional support to ensure effective ESG reporting implementation at the European level [11].

The figure below presents the ESG reporting rates among N100 (top 100 companies by revenue in each country included in the survey) companies in Europe for the years 2022 and 2024, highlighting the progression in sustainability disclosure practices over time.



Figure 1. ESG Reporting Rates in Europe in 2022 and 2024 among N100 Companies (top 100 companies by revenue in each country included in the survey), Source: Prepared by the authors based on KPMG's Survey of Sustainability Reporting 2024 [11]

The analysis of ESG reporting among the top 100 national companies in Europe reveals a clear trend toward the maturation and expansion of sustainability disclosure. Most European countries exhibit ESG reporting rates exceeding 75%, with some—such as Luxembourg and Malta—reaching full compliance (100%). This suggests a rapidly advancing alignment with European regulatory requirements, particularly the CSRD.

Western and Northern European countries—such as Germany, the Netherlands, Sweden, Switzerland, and the United Kingdom—are at the forefront of this process, consistently reporting figures close to or at 100%. These outcomes reflect not only regulatory compliance but also a strategic integration of sustainability within corporate business models.

By contrast, Eastern European countries such as Romania, Slovakia, and Estonia report comparatively lower levels of ESG disclosure. Nevertheless, these countries have demonstrated positive progress since 2022, indicating a gradual convergence toward Western standards. For instance, Romania's ESG reporting rate increased from 74% to 78%, signaling both advancement and significant room for improvement.

This regional distribution underscores the persistence of an East–West gap in ESG maturity, shaped by factors such as regulatory infrastructure, corporate culture, investor pressure, and the degree of integration into international financial markets.

The evolution of ESG reporting increasingly relies on digital technologies and real-time data integration to enhance transparency, accuracy, and operational efficiency. Across Europe and globally, organizations are adopting advanced software solutions that automate data collection, analyze sustainability metrics, and facilitate compliance with regulatory frameworks such as the ESRS and CSRD, as well as international standards like the Global Reporting Initiative (GRI).

Specialized ESG reporting platforms, including Zoei ESG and Dcycle, provide integrated systems for real-time monitoring of environmental indicators, carbon footprint analysis, and automatic report generation, significantly reducing manual effort and improving data reliability [19].

Furthermore, cutting-edge tools such as ESG Reveal leverage large language models (LLMs) to extract structured data from diverse reports, enhancing the precision and efficiency of ESG disclosure analysis [18]. Empirical evidence underscores the importance of digitalization in overcoming key barriers to sustainable investment. A survey by BNP Paribas revealed that over 70% of investors perceive insufficient and inconsistent ESG data as major obstacles to green investments, highlighting the urgent need for digital solutions that ensure accurate and timely ESG reporting [16]. Moreover, artificial intelligence-driven models like Deep Search DocQA facilitate the analysis and interpretation of sustainability reports, offering actionable insights and alignment with global reporting standards [17].

These innovations empower companies to not only meet regulatory requirements but also to enhance their sustainability performance proactively.

While digitalization and mandatory CSRD compliance are accelerating in Western Europe, emerging markets like the Republic of Moldova face a different set of structural hurdles. The transition from a fiscal-driven accounting culture (NAS) to a transparency-driven ESG framework represents not just a technical change, but a systemic challenge, as explored in the following section.

4. Multidimensional barriers to ESG Reporting

Although ESG reporting has gained considerable prominence as a tool for promoting corporate accountability and sustainable development, its practical implementation continues to face substantial limitations. These challenges arise not only from the absence of a globally harmonized regulatory framework, but also from technical and structural inconsistencies that affect the quality, comparability, and reliability of disclosed information.

Moreover, variability in institutional capacity, especially between large corporations and SMEs, contributes to unequal levels of ESG integration and reporting maturity. In many cases, ESG disclosures remain fragmented, non-standardized, and insufficiently verified, thus limiting their effectiveness as instruments for informed decision-making by investors and stakeholders [11]. These challenges are categorized across different levels—normative, technical, procedural, analytical, perceptual, and structural-and are further illustrated in Fig. 2.

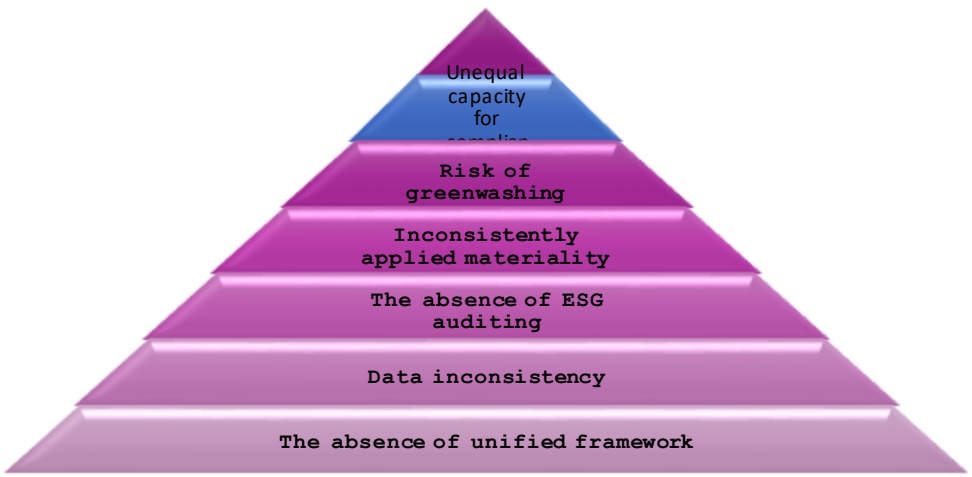


Figure 2. Gradual overview of ESG Reporting limitations
Source: Prepared by the authors, adapted from KPMG (2024) [11] and IFRS Foundation (2024) [3].

To provide further clarification of these limitations, the following section presents detailed descriptions and illustrative examples corresponding to the challenges identified in Figure 2.

- **Lack of Universal Standardization/absence of unified framework (Basic Level – Normative/Regulatory).** One of the most common limitations lies in the absence of a globally unified ESG reporting framework, resulting in fragmented application of requirements depending on the region or industry.

  Example: A multinational company may be required to comply with GRI standards in one jurisdiction, ESRS in the EU, and SASB in the US, complicating unified reporting.

- **Data Inconsistency and Quality (Technical Level – Operational).** ESG reporting frequently relies on self-declared data, without rigorous data collection methodologies or external audits. This compromises the comparability and credibility of the results [11,12].

  Example: CO_2 emissions may be estimated differently by two companies within the same sector due to the lack of standardized measurement methods.

- **The absence of ESG auditing (Procedural Level – Governance).** A significant proportion of ESG reports are not subject to external audits or other forms of independent assurance, thereby reducing confidence in their content.

  Example: Only approximately 54% of European companies that report ESG data request external verification [11].

- **Inconsistently applied materiality (Analytical Level – Strategic).** Companies often select ESG topics based on their own criteria, without rigorous analysis of the actual economic, social, or environmental impact.
  Example: Some companies report charitable actions or social campaigns but avoid addressing major environmental risks within their supply chains.

- **Risk of Greenwashing (Perceptual Level – Reputational/Ethical).** Without clear standards and external verification, some entities may use ESG reporting for purely image-related purposes, portraying exaggerated or incomplete sustainability performance.
  Example: An oil company may publish an ESG report emphasizing internal recycling efforts but omit crucial data concerning industrial pollution.

- **Unequal Capacity for Compliance (Structural Level – Institutional Development).** There is a significant gap between large companies, which have dedicated resources, and SMEs that often lack the tools, expertise, or know-how to implement complex ESG reporting [14].
  Example: In Eastern Europe, many medium-sized companies lack internal ESG teams and specialized consultants.

The gradual presentation of ESG reporting limitations reveals that the obstacles are not solely technical or legal in nature but also involve issues related to governance, ethics, and institutional capacity. An effective approach to addressing these limitations requires not only the harmonization of standards but also the strengthening of organizational culture around sustainability.

The tiered limitations described above demonstrate that the obstacles to effective ESG integration are systemic. For an emerging economy like the Republic of Moldova, these global challenges are further amplified by local institutional voids, as explored in the subsequent analysis of the national regulatory context.

5. ESG integration in the national regulatory and economic context

While ESG reporting has gained momentum globally, its development at the national level remains uneven, particularly in emerging economies. In the case of Moldova, ESG practices are still at an early stage, shaped by limited regulatory pressure, low awareness among market participants, and a lack of standardized reporting infrastructure. However, recent alignment efforts with European Union sustainability regulations—especially within the broader context of Moldova's EU integration agenda—have initiated a shift toward more structured environmental, social, and governance disclosure practices. This section explores the current ESG landscape within the country, analyzing legal developments, institutional readiness, reporting trends, and the main challenges faced by local entities in adopting ESG standards. It also considers the extent to which Moldova aligns with international best practices and what steps are needed to strengthen ESG reporting as a lever for sustainable economic and social development.

To facilitate the transition towards a sustainable economy, the Republic of Moldova has recently adopted a series of strategic policies that establish the foundational framework for national ESG integration. These governance trends signal a clear shift towards European alignment:

✓ **Environmental Strategy 2024–2030:** Approved in June 2024, this strategy targets a significant 70% reduction in net greenhouse gas (GHG) emissions by 2030 compared to 1990 levels. It focuses on cross-sectoral climate adaptation and pollution reduction, providing the 'Environmental' (E) targets that corporations will eventually need to report against [5].

To understand the scale of the reporting requirements facing national enterprises, it is essential to analyze the specific decarbonization targets set by the government. These targets, categorized by economic sector, represent the key performance indicators (KPIs) that will form the backbone of future environmental disclosures in Moldova as presented in Table 3.

Table 3. Sectoral GHG emission reduction targets by 2030 (Republic of Moldova)

Sector	GHG Emission Reduction Target (by 2030)
Energy	74% (Unconditional)
Transport	52%
Agriculture	44%
Industry	27% (Unconditional)
Waste Management	14%

Source: Prepared by the authors based on the Environmental Strategy 2024-2030 [5]

The ambitious targets presented in table 3, particularly in the energy and industry sectors, necessitate a sophisticated accounting infrastructure to track and verify progress. This highlights the critical role of implementing IFRS or aligning NAS with the requirements of Directive (EU) 2022/2464(CSRD)[4]: while National Accounting Standards (NAS) are insufficient for capturing the complex environmental liabilities and carbon-related risks associated with these targets, IFRS provides the 'fair value' and 'materiality' frameworks required for such high-level disclosures.

✓ **National GHG Monitoring and Reporting System:** By updating the mandate of the Environmental Agency, the Executive has aligned national emission inventories with the Paris Agreement and Energy Community obligations. This creates the technical infrastructure necessary for entities to access reliable secondary data for their own sustainability reports [6].

✓ **Sustainable Public Procurement (EU4Environment):** A strategic initiative aiming for at least 15% of public procurement to meet sustainability criteria by 2026. This exerts indirect pressure on the private sector to adopt ESG standards to remain eligible for state contracts [7].

✓ **Corporate Governance Reforms:** The legislative project voted in December 2024 aims to harmonize national corporate practices with EU standards, emphasizing transparency and board accountability. The **Model Code of Corporate Governance** (ratified by Government Decision No. 820/2023) establishes standardized practices for transparency, internal control, and shareholder rights for both state-owned and private enterprises, addressing the 'Governance' (G) pillar of ESG [8].

The primary barrier to effective ESG integration in Moldova is the existence of "institutional voids"—structural inefficiencies, limited institutional capacities, and regulatory ambiguity. Key organizations, such as the Environmental Agency and the Inspectorate for Environmental Protection, face constraints including insufficient personnel and overlapping mandates, which hinder the enforcement of environmental regulations and the conduct of thorough impact assessments[1,2,5]. Furthermore, as of late 2025, many administrative procedures remain paper-based, with only a small fraction of customs and environmental authorizations operational in electronic format. Bridging this digital gap is an essential prerequisite for the "data-driven institutionalization" of ESG required by the CSRD.

For SMEs in Moldova, which operate with limited resources and face weak regulatory pressure, the adoption of ESG is often driven by visionary leadership. These leaders reframe institutional gaps as strategic opportunities, using role-modeling and individualized support to embed sustainability into daily operations. By focusing on data visibility and accountability, these firms can evolve from minimal compliance to substantive governance, thereby reducing the risk of greenwashing and increasing their competitiveness in global supply chains.

Although the NAS are based on IFRS and oriented toward external stakeholders, they remain focused on financial performance. This creates a structural limitation when attempting to integrate ESG metrics, which require a broader non-financial reporting framework. In contrast, ESG reporting is value-driven, targeting a broad spectrum of stakeholders (investors, creditors, and civil society). The transition to ESG reporting is intrinsically linked to the accounting framework used by the entity. In the Republic of Moldova, IFRS serves as the 'gateway' to ESG disclosure; its principle-based nature and focus on fair value and transparency provide the necessary data architecture for sustainability reporting. Conversely, the National Accounting Standards (NAS) are not yet aligned with the non-financial disclosure mechanisms required by the CSRD, creating a significant 'reporting vacuum' for the majority of local entities.

While NAS is designed to capture quantitative, monetary-based transactions, ESG reporting requires the collection of granular, non-financial metrics (e.g., greenhouse gas emissions, social impact indicators, and board diversity). Most Moldovan enterprises lack the digital maturity and specialized software (such as integrated ERP systems) necessary to track these data points in real-time. This technological asymmetry increases the cost of compliance, particularly for Small and Medium Enterprises (SMEs) that form the backbone of the national economy.

The Moldovan accounting profession is predominantly trained in rule-based reporting (NAS), which offers limited room for professional judgment. However, international standards like IFRS S1 and S2 are principle-based and require a high degree of subjective assessment regarding materiality and climate-related risks. There is a significant shortage of locally certified ESG auditors and sustainability officers capable of translating environmental risks into financial language, creating a human capital bottleneck in the transition process.

A major limitation in the Moldovan landscape is the absence of independent assurance for sustainability disclosures. While systemic leaders (e.g., the banking sector) have begun voluntary reporting, these documents are rarely subjected to rigorous external audits. Without a national framework for ESG assurance, local reports face the risk of being perceived as "greenwashing", which undermines their utility for international green funds and institutional investors.

Moldova's banking sector, exemplified by institutions like MAIB S.A., has taken significant steps toward ESG integration. Influenced by the increasing demand for ESG-oriented assets in Europe, these institutions are beginning to monitor debt sustainability and increase the transparency of their balance sheets. This shift is not only a response to international investment pressure but a strategic move to lower equity capital costs and enhance reputation in a volatile regional market [9].

Moldcell S.A. represents the "Social" dimension of ESG leadership in Moldova. Through its commitment to digital inclusion and social impact, the company aligns its business model with the needs of the Moldovan population while adhering to international corporate governance standards. This alignment fosters stakeholder trust and enhances the company's resilience during periods of economic disruption [10].

Purcari Wineries PLC serves as a benchmark for ESG compliance within the Moldovan manufacturing and agricultural sectors. As a company listed on the Bucharest Stock Exchange (BVB), Purcari is legally mandated to adhere to rigorous transparency requirements, including IFRS-compliant financial statements and comprehensive non-financial disclosures. Their reporting framework goes beyond local requirements, integrating climate-related risks (TCFD) and water management metrics—critical factors for the viticulture industry [22]. This level of transparency is essential for maintaining investor confidence on an international regulated market and illustrates how cross-border listing acts as a catalyst for the rapid adoption of EU sustainability directives (CSRD) even for entities with core operations in Moldova.

The integration of ESG reporting within the Republic of Moldova is not merely a technical adjustment but a paradigm shift that faces significant institutional and structural barriers. The transition from the National Accounting Standards (NAS) — which are historically conservative and fiscal-oriented — to the investor-oriented IFRS/ESG framework reveals four critical gaps: regulatory, institutional, digital and market.

To bridge these gaps and ensure alignment with European sustainability trends, the transition should be structured across the following four strategic pillars:

- *Phase 1: Institutional alignment and phased mandate (ESRS based).* The Government, through the Ministry of Finance, should establish a clear timeline for mandatory ESG disclosure., aligned with the ESRS framework developed by EFRAG**.**

Priority: Start with Public Interest Entities (PIEs) that already report under IFRS (banks, insurance entities, and listed entities). For these entities, the implementation should follow the full ESRS or the Simplified ESRS for listed entities (LSME), as provided by EFRAG, to ensure international comparability [24].

Action: Transpose the EU's CSRD into national legislation, utilizing the EFRAG templates as the technical basis for reporting. This approach should allow a 2-3 year grace period for large non-listed companies currently using NAS, facilitating a gradual transition from financial-only reporting to integrated sustainability disclosures.

- *Phase 2: Development of "ESG-Lite" for SMEs.* Given that SMEs represent over 90% of the Moldovan economy, the complexity of full ESRS remains a prohibitive barrier.

Action: Moldova should adopt the VSME (Voluntary SME) framework, as developed by EFRAG[14], to serve as a simplified reporting template for non-listed entities. This "ESG-

Lite" approach focuses on essential metrics—such as energy efficiency, labor safety, and basic governance—allowing Moldovan SMEs to maintain their position within European supply chains.

As previously analyzed in the context of institutional transformations, this model allows smaller businesses to enhance transparency without facing excessive administrative burdens[25].

- *Phase 3: Capacity Building and Academic Integration.* To address the professional expertise gap, a systemic educational effort is required.

Action: Academic institutions, such as the Academy of Economic Studies of Moldova (ASEM), should integrate specialized ESG modules into accounting and management curricula, following the strategic framework for modernizing higher accounting education [23,p.7-11]. Furthermore, professional bodies (ACAP, etc.) should launch certification programs for "Sustainability Officers" to build a local pool of experts.

- *Phase 4: Digitalization and Centralized Data Governance.* To reduce information asymmetry and attract green funds, Moldova needs a transparent data infrastructure.

Action: Create a National ESG Digital Portal, managed by the National Bureau of Statistics or a dedicated agency. This platform would allow companies to upload standardized ESG data (using XBRL format), providing international investors with a "single source of truth" regarding Moldova's sustainability performance.

To overcome these barriers, the Republic of Moldova must implement an "adaptive ESG reporting model" that reduces the compliance burden for small players while providing the financial incentives, advisory services, and education needed for successful integration.

For Moldovan entities, particularly SMEs, the transition to ESG reporting involves significant compliance costs and barriers that differ from those faced by developed-market firms.

6. Conclusion

In conclusion, bridging the identified structural and institutional gaps requires a stratified approach that prioritizes the adoption of EFRAG standards (LSME and VSME) as the technical foundation for alignment with European directives. By leveraging the reporting expertise of IFRS-adopting entities and integrating sustainability into both academic curricula and digital infrastructure, Moldova can transform its current 'reporting vacuum' into a transparent ecosystem capable of attracting green capital and ensuring long-term economic resilience.

While the Moldovan ESG landscape is currently in its nascent stage, the rapid alignment with the *acquis communautaire* and the proactive stance of systemic leaders — such as MAIB, Purcari Wineries and Moldcell — provide a credible foundation for institutional change. However, the transition remains fragile, caught between the rigid legacies of fiscal-oriented accounting (NAS) and the transparent, forward-looking demands of international markets (IFRS/ESG). The proposed Roadmap offers a pragmatic path to bridge this 'implementation gap' by prioritizing institutional alignment and digital integration.

Ultimately, the success of this paradigm shift will depend on the government's ability to transform ESG from a 'check-the-box' compliance exercise into a core driver of national economic resilience and global competitiveness.

References

1. Spre o transformare verde a Republicii Moldova. Raport Naţional bazat pe setul OCDE de indicatori ai creşterii verzi [Towards a Green Transformation of the Republic of Moldova. National Report

based on the OECD set of green growth indicators], 2021. Available: https://euneighbourseast.eu/wp-content/uploads/2022/02/report-green-growth-indicators-republic-moldova-2021-rom.pdf
2. Raport național bazat pe setul OECD de indicatori ai creșterii verzi, [National Report based on the OECD set of green growth indicators], 2022. Available: (https://mediu.gov.md/indicatori/capitol0.html)
3. IFRS Foundation, 2024, New report shows global progress in corporate climate-related disclosures. Retrieved from :https://www.ifrs.org/news-and-events/news/2024/11/new-report-global-progress-corporate-climate-related-disclosures/
4. Directive (EU) 2022/2464 of the European Parliament and of the Council of 14 December 2022,(https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32022L2464)
5. Hotarare Nr. 409 din 12-06-2024 cu privire la aprobarea Strategiei de mediu pentru anii 2024-2030, [Government Decision No. 409 of 12-06-2024 on the approval of the Environmental Strategy for the years 2024-2030]. Available: (https://www.legis.md/cautare/getResults?doc_id=144295&lang=ro)
6. Press Release, 2025, (https://www.am.gov.md/ro/content/agen%C8%9Bia-de-mediu-participat-la-atelierul-regional-privind-%C3%AEmbun%C4%83t%C4%83%C8%9Birea-registrelor-de-0
7. Press Release, 2023, (https://www.mf.gov.md/ro/content/programul-eu4environment-sprijin%C4%83-republica-moldova-%C3%AEn-orientarea-cheltuielilor-publice?utm)
8. Codul-model de guvernanță corporativă al SA și SRL, publicat 2023, [Model Code of Corporate Governance for JSCs and LLCs]. Available: (https://www.bizlaw.md/public/codul-model-de-guvernanta-corporativa-al-sa-si-srl-publicat?utm_)
9. Sustainability Report 2024 MAIB, (https://www.maib.md/ro/noutati/maib-publica-raportul-de-sustenabilitate-pentru-anul-2024?utm_)
10. Management and Sustainability Report 2024, Moldcell, https://moldcell.md/files/Policies/Politici/Raportul%20de%20Management%20si%20Sustenabilitate%20pentru%20Moldcell%20Grup%20%C8%99i%20individuale%20pentru%20Moldcell%202024.pdf?utm;
11. Survey of Sustainability Reporting 2024 KMPG, 2024,(https://kpmg.com/ee/en/insights/2024/11/the-move-to-mandatory-reporting.html)
12. The CSRD Timeline: A Practical Overview, Deloitte, 2022, (https://www.deloitte.com/content/dam/assets-zone2/gr/en/docs/services/risk-advisory/2024/gr_deloitte_sustainability_eu_csrd_timeline_june_2022_noexp.pdf?utm)
13. Corporate Sustainability Reporting Directive (CSRD) — Overview, (https://ec.europa.eu/info/business-economy-euro/company-reporting-and-auditing/company-reporting/corporate-sustainability-reporting_en)
14. EFRAG. (2024). Voluntary ESRS for non-listed SMEs (VSME), https://www.efrag.org/sites/default/files/sites/webpublishing/SiteAssets/VSME%20Standard.pdf
15. EU CSRD Readiness Guide, 2025, (https://www.vixio.com/insights/pc-eu-corporate-sustainability-reporting-directive-readiness-guide)
16. BNP Paribas. (2023). Investor Survey on ESG Data Quality and Transparency. Retrieved from: https://www.ft.com/content/41125a1b-f12e-4f4f-b8e1-c11ca1fce952
17. Kim, H., Lee, S., & Park, J. (2023). Deep Search DocQA: AI Models for Sustainability Reporting Analysis. arXiv preprint arXiv:2305.08711; https://arxiv.org/abs/2305.08711
18. Zhang, X., Li, Y., & Wang, M. (2023). ESGReveal: Leveraging Large Language Models for Structured Data Extraction in ESG Reporting. arXiv preprint arXiv:2312.17264; https://arxiv.org/abs/2312.17264
19. Zoei ESG. (2024). ESG Reporting Software in Europe: Features and Benefits. Retrieved from: https://zoeiesg.com/esg-reporting-software-europe/
20. Rödl & Partner, 2026. Sustainability Reporting 2026: Obligations, Deadlines and Recommendations for all Company Sizes. Available at: https://www.roedl.com/en/insights/sustainability-reporting-2026-obligations-deadlines-and-recommendations-for-all-company-sizes/,
21. Directive (EU) 2025/794 of the European Parliament and of the Council of 14 April 2025 amending Directives (EU) 2022/2464 and (EU) 2024/1760 as regards the dates from which Member States are to apply certain corporate sustainability reporting and due diligence requirements (Text with EEA relevance), https://eur-lex.europa.eu/eli/dir/2025/794/oj/eng

22. Sustainability Report 2024, Purcari Wineries Public Company Limited. Available at: https://purcariwineries.com/rapoarte-anuale/
23. Lazari Liliana, Grigoroi Lilia. O abordare strategică pentru educația contabilă universitară în contextul imperativelor de sustenabilitate și inteligență artificială. In*: În memoriam Profesorului Viorel ȚURCANU, "Învățământul superior contabil: provocări și soluții", ediția IV, Colocviu Științific cu participare internațională*, 14 noiembrie 2025, Chișinău. ASEM. pp.7-11. https://doi.org/10.53486/isc2025.01, https://caae.ase.md/files/caae/conf/Culegere_teze_Colocviu_2025.pdf
24. EFRAG. (2025). Draft ESRS for Listed SMEs (LSME). European Financial Reporting Advisory Group. Available at: https://www.efrag.org/en/projects/draft-esrs-for-listed-smes/
25. Iavorschi Irina, Lazari Liliana. Sustainability reporting – a strategic and regulatory challenge. În: VIII All-Ukrainian Scientific and Practical Conference "Accounting, Control, and Analysis under Conditions of Institutional Changes" organized by the Department of Accounting and Taxation of Poltava State Agrarian University, Ukraine, 30 octombrie 2025. Pp. 230-233. 0,25 c.a. УДК 657. Збірник наукових праць VIII Всеукраїнської науково-практичної конференції (30 жовтня 2025 р., м. Полтава) / Полтавський державний аграрний університет. Полтава: ПДАУ, 2025. Т. 2. 939 с.

Stochastic Modeling of Cyber Resilience in Smart Power Grids Using Continuous-Time Markov Chains

Svetlana Konyuk, Alexey Rusanov

S.C. "Add-Production" S.R.L., Chisinau, Moldova
Svetlana.Coniuc@gmail.com, Rusanov.Alexei@gmail.com

Abstract. This article introduces a stochastic model for evaluating the resilience of smart power grids against cyberattacks aimed at distorting data flows. The model is grounded in continuous-time Markov chains (CTMC) and captures the probabilistic transitions of the system from normal operation to states of local degradation, large-scale compromise, and ultimately functional failure. The analysis provides a comprehensive treatment of attack propagation mechanisms, including the compromise of meters and concentrators, telemetry substitution, and the exploitation of asynchronous notifications, thereby enabling the representation of both direct and cascading effects. The proposed framework delivers a quantitative assessment of incident likelihood, recovery time, and expected impact, with explicit consideration of metrics such as Data Integrity Loss (DIL) and analogues of well-established energy reliability indices, including Expected Unserved Energy (EUE) and Loss of Load Probability (LOLP). Particular emphasis is placed on the economic dimension: the modeling results demonstrate the existence of a critical investment range in cybersecurity, within which the return on investment is maximized while costs remain minimal. The practical significance of this work lies in its potential application to the development of corporate information security policies, the adaptation of national regulatory standards, and the design of decision support systems for the energy sector.

Keywords. Cyber resilience, smart grids, continuous-time Markov chains, risk assessment, information security.

1. Introduction

The development of smart power grids (Smart Grids) represents a key direction in the digital transformation of the energy sector. These systems integrate distributed generation sources, including renewables, and enable bidirectional interaction with consumers. Unlike traditional power systems, the Smart Grid functions as a cyber-physical system in which energy transmission processes are closely coupled with information flows. This integration provides opportunities for improving energy efficiency and infrastructure reliability, while at the same time increasing vulnerability to cyber threats.

Ensuring the information security of smart power grids has become a strategic priority, as disruptions in the operation of their digital components can trigger large-scale supply failures and cause substantial socio-economic damage. In the context of the growing number of cyberattacks targeting critical infrastructure, the development of methods to enhance the cyber resilience of the Smart Grid is acquiring particular urgency.

The objective of this study is to develop a stochastic modeling framework for assessing the cyber resilience of smart power grids, providing a foundation for both risk evaluation and the formulation of protective strategies.

To achieve this objective, the following tasks are to be addressed:

- Conduct an analysis of threats and vulnerabilities of cyber-physical components in the Smart Grid.
- Examine existing international standards and best practices for the protection of critical infrastructure.
- Develop a methodology for assessing the cyber resilience of the power grid.
- Perform validation of the proposed approach using model scenarios or an experimental testbed.

The scientific contribution of this research lies in the development of an integrated probabilistic framework for ensuring the cyber resilience of smart power grids, which combines probabilistic mathematical modeling, risk-oriented analysis, and alignment with established information security standards. In contrast to the majority of existing studies, the proposed approach goes beyond mere threat identification based on a continuous-time Markov chain (CTMC) model that explicitly captures the interdependence between energy and information subsystems and links attack dynamics to operational reliability metrics (DIL, MTTD, MTTR, EUE-analogue).

The relevance of the study is shaped by the convergence of two pressing factors: the urgent need to digitalize the energy sector to enhance efficiency and sustainability, and the escalating risks of cyberattacks capable of undermining the very advantages of Smart Grid deployment. To bridge this gap, the present research contributes an interdisciplinary methodology that unites energy engineering with information technology, offering both theoretical advancement and practical value for critical infrastructure protection.

2. Analysis of Critical Data Flows and System Architecture

The conducted analysis of data flows and system architecture has revealed several critical points in the "meter–concentrator–data center" chain that require increased attention from the perspective of both information and technical security.

The most vulnerable components are the communication channels that employ power line communication (PLC) and wireless technologies with insufficient protection. Interception or substitution of data transmitted through these channels can result in widespread falsification of measurements, disruption of load-balancing processes, and even local power supply interruptions.

At the communication layer, which is responsible for aggregating and balancing distributed generation, security breaches may prevent real-time data on electricity production and consumption from reaching the control center. Such disruptions can lead to mismatches between generation and demand, an increased number of emergency situations, and reduced stability of the entire power system.

Asynchronous notification mechanisms are designed to transmit critical events to the data center, such as alerts on equipment failures or signs of tampering. However, once compromised, these channels may be exploited by adversaries to inject false signals. Consequently, disinformation attacks can be launched, provoking incorrect responses from operators or automated control systems [6]-[10].

3. Resilience Model of the Power Grid under Cyberattacks

The proposed model describes the probabilistic dynamics of a power grid under cyberattacks targeting its information flows. Particular attention is given to the system's transition from normal operation to states of degradation and, ultimately, to functional failure ("data-driven blackout") caused by the loss of data integrity.

The model takes into account the following factors:

- the frequency of attacks on communication channels and data aggregation nodes;
- the probability of successful compromise of meters, concentrators, or the control center;
- the propagation of attacks within the network, including false event injection and cascading errors;
- processes of detection, response, and recovery (such as patching, switching to backup channels, and equipment replacement).

For formalization, the model employs a continuous-time Markov chain (CTMC) [1]-[5] with aggregated states. When necessary, the framework can be extended to semi-Markov or phase-type models in order to capture non-exponential detection, response, and recovery times associated with cyberattacks.

4. Model States

Within the framework of the proposed model, four states are distinguished:

S_0 – Normal Operation. All data remain reliable, and control processes are performed correctly.

S_1 – Local Compromise. A subset of devices is compromised; data are partially distorted, and restrictive measures are applied.

S_2 – Large-Scale Compromise. Telemetry substitution and false notifications become systemic, creating risks of incorrect control commands and mismatches between generation and load.

S_3 – Functional Failure. Control becomes impossible due to the loss of data integrity; the consequences are comparable to a blackout.

State Transitions

The transitions of the system from one state to another are described by transition rates (measured in units of 1/hour), which reflect the probability that a change of state will occur within a short time interval. These transition rates enable the formalization of the process using continuous-time Markov chains and provide a means to link the system's technical parameters with the probabilities of incidents.

Compromise Transitions

$S_0 \rightarrow S_1$:

$a_{01}=\lambda p c_a$, where λ is the attack frequency and p_c is the probability of a successful attack. This rate represents the likelihood that, given a certain attack frequency (λ) and probability of success (p_c), the system transitions from normal operation to local degradation. In essence, the more frequent the attacks and the higher the success rate, the faster individual devices (meters or concentrators) are compromised.

$S_1 \rightarrow S_2$:

$a_{12}=\lambda p_c \cdot k_{12}$, where k_{12} reflects the complexity of attack propagation. At this stage, the attack escalates from a local issue to a systemic compromise. The coefficient k_{12} reduces the transition rate, reflecting the additional effort required for an adversary to propagate the attack (e.g., spreading from one concentrator to adjacent nodes).

$S_2 \rightarrow S_3$:

$a_{23}=\lambda p_c \cdot k_{23}$. This stage corresponds to a complete informational collapse: control loses reliability, and the risk of a "functional blackout" emerges. The coefficient k_{23} is typically smaller than k_{12}, since achieving a full-scale system failure is more difficult than compromising a portion of the infrastructure.

Internal Propagation (Cascade Effects)

$S_1 \rightarrow S_2 : \gamma_1$

Even if the attack does not escalate directly through new compromise attempts, the system may deteriorate due to internal cascading effects. For example:

- an adversary substitutes several meter messages,
- the concentrator interprets them as legitimate problems,
- the control center reacts incorrectly, thereby amplifying the degradation.

The parameter γ_1 reflects precisely such "self-developing" errors. It represents the likelihood that a local disruption evolves into a systemic compromise due to the internal dynamics of the network.

$S_2 \rightarrow S_3 : \gamma_2$.

At this stage the situation becomes even more severe: an existing large-scale compromise may, by itself, lead to a complete loss of controllability. For instance:

- false alarm events overload operators,
- automatic control systems receive contradictory data and execute erroneous switching actions,
- as a result, the system experiences an abrupt transition to a "functional blackout."

Thus, the parameter γ_2 represents the risk that, even in the absence of new external attacks, cascading internal failures can drive the network into total collapse.

Detection and Recovery

$S_1 \rightarrow S_0 : \mu_1$.

This parameter corresponds to rapid recovery measures, where local compromise (e.g., an infected meter or concentrator) is promptly detected and isolated. In such cases, administrators can disconnect the compromised device, clean it, or roll back its configuration. The parameter μ_1 represents the recovery rate, i.e. the intensity of returning from S_1 to S_0. A higher value of μ_1 corresponds to a shorter mean recovery time, thereby reducing the likelihood that a local attack will escalate into a systemic compromise.

$S_2 \rightarrow S_1 : \mu_2$.

This case involves more complex incidents in which the compromise has already spread further. Recovery requires substantial measures such as reissuing access keys, replacing firmware or hardware, or re-establishing trusted communication channels. The recovery rate μ_2 is typically lower than μ_1, reflecting longer mean recovery times. It formally quantifies the organization's ability to cope with severe cyberattacks.

$S_3 \rightarrow S_1 : \rho_1$.

Recovery from the functional failure state S_3 may proceed through direct transitions to to S_0 or S_1; however, staged recovery via S_2 is explicitly relevant and can be integrated into the model to better reflect real-world processes.

Parameter Semantics (CTMC Rates)

In a continuous-time Markov chain, each off-diagonal element q_{ij} is a transition rate (hazard). For state S_i, the mean sojourn time equals $1/(-q_{ii})$. Accordingly, μ_1, μ_2, ρ_0, and ρ_1 are rates of recovery transitions:

μ_1 – rate from S_1 to S_0 (local recovery),
μ_2 – rate from S_2 to S_1 (de-escalation),
ρ_1 – rate from S_3 to S_1 (emergency partial restoration),
ρ_0 – rate from S_3 to S_0 (full restoration).

Additionally, δ denotes the detection rate. It is explicitly included in the generator matrix Q and directly contributes to the recovery transitions ($S_1 \rightarrow S_0$ and $S_2 \rightarrow S_1$), thereby ensuring transparency and completeness of the model

If one wishes to discuss times rather than rates, the corresponding mean recovery times are $1/\mu_1$, $1/\mu_2$, $1/\rho_1$, and $1/\rho_0$.

The dynamics are governed by the CTMC generator Q where each diagonal element is defined as $q_{ii} = -\sum_{i \neq j} q_{ij}$, ensuring that the row sums equal zero. The explicit form of the generator is as follows (1):

$$Q = [-a_{01}\ a_{01}00\ \mu_1 - (\mu_1 + a_{12} + \gamma_1)\ a_{12} + \gamma_1 00\ \mu_2 - \\ -(\mu_2 + a_{23} + \gamma_2)\ a_{23} + \gamma_2\ \rho_0 \rho_1 0 - (\rho_0 + \rho_1)] \quad (1)$$

Here:

For state S_0: $q_{00} = -a_{01} q_{01} = a_{01}$.
For state S_1: $q_{11} = -(\mu_1 + a_{12} + \gamma_1)$, $q_{10} = \mu_1$, $q_{12} = a_{12} + \gamma_1$.
For state S_2: $q_{22} = -(\mu_2 + a_{23} + \gamma_2)$, $q_{21} = \mu_2$, $q_{23} = a_{23} + \gamma_2$.
For state S_3: $q_{33} = -(\rho_0 + \rho_1)$, $q_{30} = \rho_0$, $q_{31} = \rho_1$.

This formulation ensures that each row of Q sums to zero, thereby satisfying the fundamental property of CTMC generators.

Initial distribution: $p(0)=[1,0,0,0]$.
Time evolution: $p(t)=p(0)e^{Qt}$.
Stationary distribution: $\pi Q=0$, $\sum \pi_i = 1$.

Based on the model, one can compute:

- the probability of functional failure by time t (2):

$$P_{BL}(t)=p_3(t) \text{ (the } S_3 \text{ component).} \quad (2)$$

- the **Data Integrity Loss (DIL)** — the expected fraction of unreliable data;
- the mean time to detection (MTTD) and mean time to recovery (MTTR);
- the expected impact of distorted data on power supply (an analogue of **Expected Unserved Energy**, EUE).

Metrics and Evaluation Procedures (DIL, MTTD, MTTR, EUE-analogue)

The proposed CTMC-based model allows the derivation of quantitative resilience indicators that capture both reliability and security aspects of smart grid operation under cyberattacks.

This metric measures the expected share of corrupted or unreliable data. It is defined as the weighted probability of being in a compromised state (S_1–S_3) (3):

$$DIL(t)=p(t)\ w^T, \quad (3)$$

where: $p(t)=[p_0(t),p_1(t),p_2(t),p_3(t)]$ is the probability distribution over states at time t, $w=[0, w_1, w_2, w3]$ is the weighting vector that specifies the relative share of unreliable data in each state. By convention, $w_0=0$ (no data corruption in the normal state), while $w_3=1$ corresponds to complete data loss in state S_3.

Typical assumptions: $w_1 \ll w_2 < 1$, reflecting that local compromise leads to minor distortion, large-scale compromise entails significant corruption, and functional failure implies total loss of data integrity.

MTTR (Mean Time to Recovery) denotes the expected time required to restore the system to a stable operational state following a compromise. In addition, for specific scenarios, a partial recovery time (pMTTR) can be defined for the subset of states $\{S_1, S_2\}$. This avoids confusion between full and partial recovery times (4):

$$T = -Q_{TT}^{-1} 1 \quad (4)$$

If a scalar MTTR is required for a given initial distribution α\alphaα over the transient states, it is computed as MTTR = αT.

Analogous to the reliability index Expected Unserved Energy (EUE), this metric reflects the expected impact of compromised data on power supply. Each state S_i is assigned an energy loss factor c_i, corresponding to the proportion of demand potentially mismanaged . The expected unserved energy is then (5):

$$EUE(t) = p(t)c^T, \quad (5)$$

where $c=[0, c_1, c_2, c_3]$. Here
c_1 reflects minor inaccuracies in load balancing,
c_2 accounts for systemic miscoordination,
$c_3 \approx 1$c corresponds to blackout-level disruption.

Interpretation:

DIL quantifies the *data-level integrity degradation*.

MTTD and **MTTR** characterize *operational responsiveness*.

EUE-analogue links the cyber perspective with *energy supply consequences*.

Together, these indicators provide a comprehensive assessment of cyber resilience by connecting **attack dynamics** (λ, p_c, γ) with operational outcomes (availability, reliability, and socio-economic impact).

5. Practical Application of the Model Analysis of Cyberattack Scenarios

The developed model makes it possible to examine in detail various cyberattack scenarios targeting the power system and to forecast their consequences.

Frequent but low-impact attacks

Such attacks do not lead to severe consequences but impose a constant load on detection and response subsystems. The model shows that when the attack frequency (λ) is high but the success probability (p_c) is low, the likelihood of the system reaching critical states (S_2–S_3) remains minimal. At the same time, the risk of overloading monitoring subsystems and generating false alarms increases.

Rare but targeted campaigns

When attacks occur infrequently but with high efficiency (high p_c), the probability of rapid escalation to large-scale compromise and functional failure rises significantly. Under such conditions, the key resilience factors are detection speed (δ) and recovery rates (μ_2, ρ).

Combined attacks (technical + social engineering).

This scenario is characterized by an increase not only in the probability of success (p_c) but also in the cascading coefficients (γ_1, γ_2). Social engineering may grant attackers access to internal systems, substantially raising the risk of escalation. The model confirms that such hybrid attacks pose the most severe threat to power grids.

Exploitation of asynchronous notifications

In this case, the parameter γ_1 rises sharply. The system may treat false events as genuine incidents, leading to cascading disruptions. Simulation results indicate that under high values of γ_1, even local compromises can evolve into systemic crises.

Conclusion: The key resilience factors across most scenarios are the speed of detection and recovery, as well as the reduction of compromise probability through regular patching and communication channel segmentation [6], [9].

6. Economic Efficiency of Investments

The model enables not only the forecasting of attack consequences but also the evaluation of the economic efficiency of various cybersecurity measures.

The analysis results revealed the existence of a critical investment range. Simulation results under assumed parameter values indicate that simulation experiments under assumed parameter values suggest the existence of a critical investment range. Illustrative simulations suggest the existence of a potential investment threshold, where increasing cybersecurity expenditures significantly reduces expected damage, while further increases show diminishing returns. The specific figures (e.g., 20% of the IT budget) are scenario-dependent and require validation with empirical data. These values are illustrative and depend strongly on the assumptions regarding attack frequency, success probability, and recovery rates.

This outcome has important practical implications:

- it supports the development of an optimal strategy for resource allocation;
- it confirms the necessity of applying methods of economic and mathematical analysis to identify priority protection areas;
- it helps to avoid both underinvestment (insufficient protection) and overspending with low marginal effectiveness.

7. Practical Significance

The proposed models can be applied in the following contexts:

- in the formation of corporate cybersecurity policies for energy companies;
- in the adaptation of national and sectoral information security standards to the conditions of energy digitalization;
- as a foundation for the development of decision support systems focused on risk management and cost optimization.

Thus, the proposed methodology combines technical and economic modeling, thereby providing a comprehensive approach to the problem of cyber resilience in energy infrastructure.

References

1. K. A. Shcheglov and A. Yu. Shcheglov, “Mathematical models of operational information security,” Information Protection Issues, vol. 106, no. 3, pp. 52–65, 2014.
2. K. A. Shcheglov and A. Yu. Shcheglov, “Operational characteristics of information system security risks,” Scientific and Technical Journal of Information Technologies, Mechanics and Optics, no. 1 (89), pp. 129–139, 2014.
3. A. P. Rozhenko, Internal Threats to the Security of Confidential Information: Methodology and Theoretical Research. Moscow: Krasand, 2010.
4. E. B. Belov, V. P. Los, R. V. Meshcheryakov, and A. A. Shelupanov, Fundamentals of Information Security. Moscow: Goryachaya Liniya—Telekom, 2006.

5. K. A. Shcheglov and A. Yu. Shcheglov, "Protection against malicious software by controlling access to created file objects," Herald of Computer and Information Technologies, no. 8, pp. 46–51, 2012.
6. Y. Yan, Y. Qian, H. Sharif, and D. Tipper, "A survey on cyber security for smart grid communications," IEEE Communications Surveys & Tutorials, vol. 14, no. 4, pp. 998–1010, 2012.
7. X. Fang, S. Misra, G. Xue, and D. Yang, "Smart grid – The new and improved power grid: A survey," IEEE Communications Surveys & Tutorials, vol. 14, no. 4, pp. 944–980, 2012.
8. A. R. Metke and R. L. Ekl, "Security technology for smart grid networks," IEEE Transactions on Smart Grid, vol. 1, no. 1, pp. 99–107, 2010.
9. H. He and J. Yan, "Cyber-physical attacks and defenses in the smart grid: A survey," IEEE Transactions on Smart Grid, vol. 7, no. 1, pp. 281–299, 2016.
10. Y. Mo, T. H.-J. Kim, K. Brancik, D. Dickinson, H. Lee, A. Perrig, and B. Sinopoli, "Cyber-physical security of a smart grid infrastructure," Proceedings of the IEEE, vol. 100, no. 1, pp. 195–209, 2012.

Investments in Moldova's Energy Infrastructure: Priorities and Modernization Prospects

Timur Lesan

Academy of Economic Studies of Moldova, Chisinau, Moldova
lesan.timur@ase.md

Abstract. The article examines investments in Moldova's energy infrastructure as a key factor for its modernization and sustainable development. The current state of the infrastructure, main directions, and investment priorities are analyzed, including the development of power grids, renewable energy sources, storage systems, and digitalization. Special attention is given to financing sources and the role of international organizations in supporting the modernization of the sector. The prospects for the development of Moldova's energy system are outlined in the context of integration into the European energy market and the transition to a low-carbon economy.

Keywords. Energy infrastructure; Energy investments; Renewable energy sources; Power grids; Energy storage systems.

1. Introduction

The modernization of energy infrastructure is a crucial factor in ensuring sustainable economic growth and energy security. For the Republic of Moldova, this issue is particularly relevant due to its high dependence on energy imports, outdated infrastructure, and the strategic objective of integration into the European energy market. More than two-thirds of Moldova's electricity demand is covered by external sources, while domestic generation capacities remain limited and are largely based on fossil fuels or facilities inherited from the Soviet period. In this context, investments in modern energy infrastructure represent both an economic necessity and an opportunity to align with the objectives of the European Union's Green Deal.

The purpose of this article is to analyze the current state of Moldova's energy infrastructure, identify investment priorities, and assess modernization prospects in accordance with advanced international experience.

Current State of Moldova's Energy Infrastructure

Moldova's electricity sector is characterized by a high dependence on imports and the absence of diversified domestic generation. A significant share of electricity supply was provided by the Moldavian GRES power plant until 2025, located in the Transnistrian region, which created both economic and political vulnerabilities. Since 2025, a larger share of electricity has been supplied from Romania. Renewable energy sources (RES), mainly solar energy and biomass, account for a small but steadily growing share of the energy mix.

Transmission and distribution networks face serious challenges. According to national statistics, electricity losses in transmission and distribution are among the highest in the region, reaching 11–12%. Many power lines, substances, and transformers were built more than 30 years ago and require urgent modernization. Limited grid capacity also constrains the integration of new renewable energy projects [13].

2. Research results

As we can see, 17% of the required electricity was purchased on the exchange. During the winter months, purchases on the exchange reached 25%.

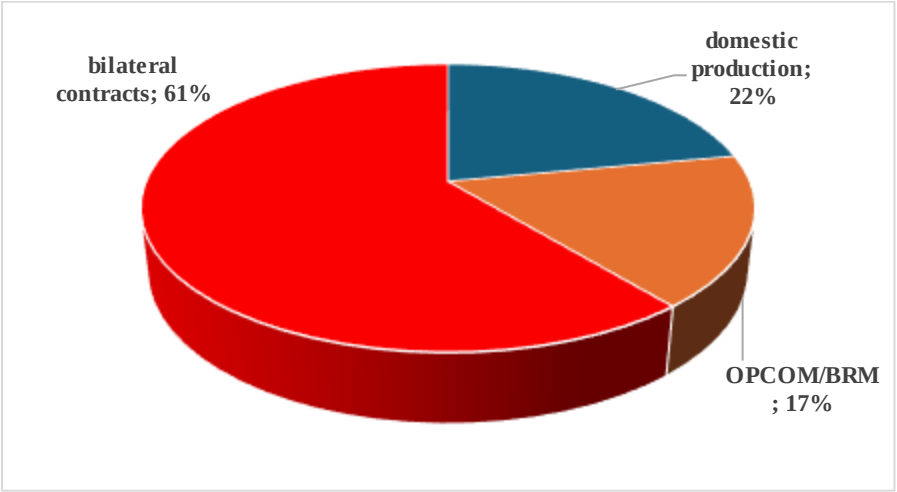


Figure 1. The current state of Moldova's energy infrastructure
Sourse: compiled by the author based on data from https Energocom.md

Table 1. Electricity consumption and generation from renewable energy sources in Moldova

Indicators	Unit	2020	2021	2022	2023	2024
Total amount of purchased electricity	million kWh	**4 269,80**	**4 591,70**	**4 512,90**	**4 333,10**	**4 626,10**
% change	%		8%	-2%	-4%	7%
Production	million kWh	**851,4**	**984,7**	**851,1**	**1 010,80**	**1 340,80**
% change	%		16%	-14%	19%	33%
Attitude to purchased energy	%	20%	21%	19%	23%	29%
of which from renewable energy source	million kWh	**129,9**	**186,8**	**239,6**	**407,3**	**643,1**
% change	%		44%	28%	70%	58%
Attitude to purchased energy	%	3%	4%	5%	9%	14%

Sourse: compiled by the author based on data from https anre.md

Electricity consumption increased slightly, while production from renewable energy sources reached 14% in 2024. Moreover, 70% of energy is imported either from Ukraine and Romania or from state district power plants.

The amount of energy entering the grid has increased in recent years. Until 2025, the majority of this energy came from the Moldavian Regional Power Plant (MGRES). Since 2025, imports from Romania have increased, as have those from renewable energy sources. In the first half of 2025, the volumes have equaled those for the entire year of 2024. We have also observed an increase in process losses in recent years.

Table 2. The amount of electricity received by the transmission network of Moldova

Indicators	2019	2020	2021	2022	2023	2024	2 кв 2025
The amount of electrical energy supplied to the transmission grid, million kWh	4294,5	4254,4	4550,6	4427,97	4617,51	5014,2	2413,8
inclusive:							
MGRES, million kWh	2856,8	3251,3	3445,6	2708,34	3293,7	3093,6	0
Thermal power plant, million kWh	659,6	721,5	797,7	611,52	603,1	698	424,2
Renewable energy, million kWh	134,1	114,5	145,8	153,02	261,51	478,3	453,4
Import, million kWh	644	167,1	161,5	955,09	459,2	744,3	1536,2
Technological losses %	2,48%	2,43%	2,56%	3,66%	4,05%	14,82%	6,90%

Sourse: compiled by the author based on data from https www.moldelectrica.md/anre.md

Table 3. Length of overhead electrical lines in Moldova

Length of overhead electrical lines	2019	2020	2021	2022	2023	2024	2 кв 2025
35 kV	807,59	807,59	807,59	807,59	807,59	668,3	668,3
110 kV	3337,04	3337,04	3337,04	3337,04	3337,04	2818,37	2818,37
330 kV	377,34	377,34	377,34	377,34	377,34	377,34	377,34
400 kV	203	203	203	203	203	203	203
Number of power transformers in operation	648	650	650	649	650	650	650
Total capacity of power transformers in operation	4750,2	4775,2	4775,2	4650,5	4641,5	4641,5	4656,5

Sourse: compiled by the author based on data from https www.moldelectrica.md

As we can see, the length of overhead lines decreased in 2024, especially between 110 kV and 519 kV. The number of power transformers remained virtually unchanged, while their capacity decreased. Given the growth in energy consumption and the variety of supplies, the need for investment in new lines and transformers is only increasing (most networks and equipment are over 30 years old).

Table 4. Investments approved by ANRE and executed

Indicators	Operator	Unit	2019	2020	2021	2022	2023	2024
Transmission System Operator	Î.S. „Moldelectrica"	thousand lei	380 392	136 881	88 280	16 753	161 540	700 964
Distribution system operators	Î.C.S. „Premier Energy Distribution" S.	thousand lei	259 710	321 136	258 142	245 599	311 486	389 648
	S.A. „RED Nord"	thousand lei	255 217	236 834	270 456	198 379	219 347	236 474
Universal service providers and providers	Î.C.S. „Premier Energy" S.R.L.	thousand lei	454	3 455	772	965	1 505	729
of last resort	S.A. „Furnizarea Energiei Electrice Nor	thousand lei	202	66	378		96	4 000
Total			**895 975**	**698 372**	**618 028**	**461 696**	**693 974**	**1 331 815**

Sourse: compiled by the author based on data from www.moldelectrica.md/anre.md

As we can see, investment in power grids declined during the pandemic and has increased significantly over the past two years, more than doubling. This is primarily due to Moldelectric.

Let's look at the structure of renewable energy capacities commissioned in Moldova at the end of July 2025

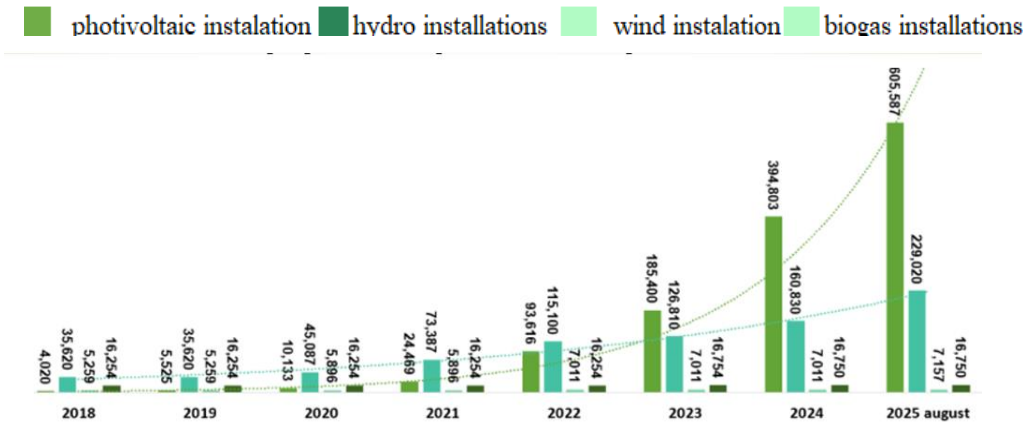


Figure 2. Investment priorities in the energy sector
Sourse: compiled by the author based on data from https:// cned.gov.md

As we can see, the growth of commissioned capacities has increased, especially solar panels and slightly less wind power plants.

	Producer					TOTAL
	Law 160-XVI/2007	Law 10/2016/fixed tariff	Free market	Net Acounting	Net invoicing	
Solar panels	0,59	139,62	241,38	115,31	48,71	545,61
Wind power plants	27,23		187,35			214,58
Biogas	5,17	1,99				7,16
Hydropower	0,25		16,5			16,75
Total	33,24	184,11	439,94	115,31	58,43	784,09

Figure 3. Capacity of renewable energy power plants commissioned at the end of July 2025
Sourse: compiled by the author based on data from https:// cned.gov.md

Compared to 2024, capacity has increased by 46%. This is primarily due to the expansion of solar panels and, to a lesser extent, wind power plants.

With the increasing deployment of new renewable energy capacity, particularly solar power plants and wind power plants, the imbalance between electricity production and consumption is widening. Solar power plants produce variable amounts of electricity depending on the season and time of day, as well as the current weather conditions.

Therefore, there is a need to improve energy transmission and storage infrastructure.

Smart grid is a digital power grid management system that enables real-time balancing of supply and demand, integration of renewable energy sources (RES), and increased energy efficiency. It utilizes sensors and smart meters. For Moldova, this means increasing grid flexibility and readiness for the growing share of RES.

Cross-border connections: lines with Romania (Vulcanesti-Isaccea, Straseni-Vulcelele). To eliminate supply risks through the Transnistrian region and reduce losses, the Vulcanesti-Chisinau line is being built. Construction of the 400 kV Vulcanesti-Chisinau

overhead power line is ongoing. Eighty percent of the supports have been installed, and over 79 km of wires have been laid, representing 51% of the line's total length (157 km).[6]

The Straseni-Gutinas power transmission line, which will connect Moldova and Romania, will require $130 million in investment.[8]

Next, another high-voltage line will be built between Balti and Suceava.

Battery parks and a study of pumped-storage hydroelectricity potential are also needed to improve infrastructure.

The installed renewable energy capacity of 720 MW at the end of July will allow for a 30% increase in consumption within 12 months, and during the summer months and with favorable weather, generation may exceed consumption.

We see increasing risks of an imbalance between production and consumption. This requires a well-developed grid interconnected with neighboring countries to transfer surplus energy and generate it when needed, or to develop energy storage systems.

Many European countries have faced such problems, forcing them to sell electricity at negative prices, limit production, or implement energy storage systems. Some European countries have been forced to limit renewable energy production.

In Europe, restrictions have increased in some places and are expected to increase further—research and analysis show that batteries can significantly reduce economic restrictions. Approximately, restrictions range from 3.5% to 10% depending on the region or country. European countries are attempting to reduce production restrictions through the introduction of energy storage systems or pumped hydroelectric power plants.

The impact of energy storage deployment varies across countries.

Spain: Modeling shows that 200 Wh batteries per 1 kW PV significantly reduce constraints; in some areas, this can be reduced by up to 80% (research estimate). According to another estimate, after the deployment of energy storage batteries, production constraints will be 0.5–1% (with a battery of ~200 Wh/kW PV).[14]

For Germany, estimates and analyst commentary have been published: batteries reduce negative price hours and contribute to grid flexibility; however, the scale of ESS is not yet sufficient to completely eliminate constraints—major capacity increases will be required. A reduction in negative hours and production constraints will occur with 2–5 GW of batteries (model estimate).[15] In Ireland, after the introduction of energy storage batteries, renewable energy generation constraints will decrease to 1–3% (depending on the battery)—empirical data.[16]

3. Prospects for modernization

Financing options:

International financial institutions (EBRD, EIB, World Bank, USAID, EU grants);

Attracting private capital through guarantees and tariff policy.

Country experiences:

Romania – rapid growth of RES with the support of green certificates;

North Macedonia – successful attraction of FDI to solar parks;

Bulgaria – problems with tariff policy as a lesson for Moldova.

4. Barriers and challenges

Financial: limited internal capital, high borrowing costs.

Regulatory: inconsistent tariff policy, unstable laws.

Technical: worn-out networks, lack of storage facilities.

Social: low consumer solvency, resistance to tariff increases.

5. Conclusion

Investments in energy infrastructure are key to sustainable economic growth.

Priority areas include energy storage, grid modernization, energy storage systems, and interconnectors.

Recommended actions:

Create a stable and transparent investment environment;

Strengthen partnerships with international organizations;

Implement European tariff incentive practices;

Develop public-private partnership mechanisms.

This will also lead to:

Increased GDP due to increased investment.

Job creation in the energy and construction sectors.

Reduced dependence on energy imports and increased energy security.

Increased competitiveness of enterprises due to reduced energy costs.

Improved environmental conditions (reduced carbon dioxide emissions).

References

1. Moldelectrica. (n.d.). Annual Report. Retrieved from https://www.moldelectrica.md/ru/network/annual_report
2. National Agency for Energy Regulation (ANRE). (2023). Report on the Activity of ANRE in 2022. Retrieved from https://anre.md/storage/upload/administration/reports/1138/Raport%20privind%20Activitatea%20ANRE%20in%20anul%202022.pdf
3. National Agency for Energy Regulation (ANRE). (2024). Annual Activity Report 2023. Retrieved from https://www.anre.md/storage/upload/administration/reports/1319/Raportul%20de%20activitate%202023.pdf
4. World Bank. (2021). Moldova Electric Power Market Options Sector Study. Retrieved from https://documents1.worldbank.org/curated/en/195111631903776975/pdf/Moldova-electric-power-market-options-sector-study.pdf
5. National Center for Sustainable Energy (CNED). (n.d.). Capacități instalate. Retrieved from https://cned.gov.md/ro/content/capacitati-instalate
6. Moldpres. (2024). Okolo 80 opor LEP Vulkaneshty-Kishinev byli ustanovleny. Retrieved from https://www.moldpres.md/rus/ekonomika/okolo-80-opor-lep-vulkaneshty-kishinev-byli-ustanovleny
7. Energocom. (2025). Achizițiile de energie electrică în luna septembrie. Retrieved from https://energocom.md/achizitiile-de-energie-electrica-in-luna-septembrie/
8. Moldova 1. (2025). Rechan: liniya elektroperedachi Strasheni Gutinash planirat da dadat v eksploatatsiya do 2028 godina, Retrieved from https://moldova1.md/p/57474/%D1%80%D0%B5%D1%87%D0%B0%D0%BD-%D0%BB%D0%B8%D0%BD%D0%B8%D1%8E-%D1%8D%D0%BB%D0%B5%D0%BA%D1%82%D1%80%D0%BE%D0%BF%D0%B5%D1%80%D0%B5%D0%B4%D0%B0%D1%87%D0%B8-%D1%81%D1%82%D1%80%D0%B0%D1%88%D0%B5%D0%BD%D1%8B-%D0%B3%D1%83%D1%82%D0%B8%D0%BD%D0%B0%D1%88-%D0%BF%D0%BB%D0%B0%D0%BD%D0%B8%D1%80%D1%83%D1%8E%D1%82-

%D1%81%D0%B4%D0%B0%D1%82%D1%8C-%D0%B2-%D1%8D%D0%BA%D1%81%D0%BF%D0%BB%D1%83%D0%B0%D1%82%D0%B0%D1%86%D0%B8%D1%8E-%D0%B4%D0%BE-2028-%D0%B3%D0%BE%D0%B4%D0%B0

9. Biroul Național de Statistică al Republicii Moldova (National Bureau of Statistics of the Republic of Moldova). (2026). Energetica – Indicatori cheie. Retrieved from https://statistica.gov.md/ro/statistic_indicator_details/14#about
10. International Finance Corporation (IFC). (2023). Creating Markets in Moldova: Country Private Sector Diagnostic. Retrieved from https://www.ifc.org/content/dam/ifc/doc/2023-delta/cpsd-moldova-en.pdf
11. Ministry of Energy of the Republic of Moldova. (2025, July 4). 30 Grid Connection Permits with a Total Capacity of Over 109 MW Returned to Electricity Transmission and Distribution Operators by June 30 of This Year. Retrieved from https://energie.gov.md/en/content/30-grid-connection-permits-total-capacity-over-109-mw-returned-electricity-transmission-and
12. National Agency for Energy Regulation (ANRE). (2021). Raport anual de activitate 2020. Retrieved from https://anre.md/raport-de-activitate-3-10 (download section for Raport anual de activitate 2020)
13. Government of the Republic of Moldova. (2025). Raport ODD 2025. Retrieved from https://gov.md/sites/default/files/media/documents/2025-12/Raport%20ODD%202025_FINAL.pdf
14. Sanchez Molina, P. (2025, July 1). Batteries with 200 Wh storage per kW of solar capacity could mitigate curtailment in Spain. ESS News. Retrieved from https://www.ess-news.com/2025/07/01/batteries-with-200-wh-storage-per-kw-of-solar-capacity-could-mitigate-curtailment-in-spain/
15. Eckert, V. (2024, September 25). Germany, EU must match green power roll-out with more batteries, says Ember. Reuters. Retrieved from https://www.reuters.com/markets/europe/germany-eu-must-match-green-power-roll-out-with-more-batteries-says-ember-2024-09-25/
16. Lagutins, N. (2024). Battery Energy Storage Systems for Wind Farm Integration and Curtailment Reduction: A Case Study in Ireland. Technological University Dublin. Retrieved fromhttps://www.researchgate.net/publication/377500304_Battery_Energy_Storage_Systems_for_Wind_Farm_Integration_and_Curtailment_Reduction_A_Case_Study_in_Ireland

Mind Games and Machines: Social Engineering in the Cybersecurity Era

Bradu Nichita

IU International University of Applied Sciences, Berlin, Germany,
bradusecure@gmail.com

Abstract. Social engineering, the deliberate manipulation of human psychology to obtain information, access, or actions that would otherwise be withheld, now underpins an estimated 98% of all cyberattacks (Sonowal, 2023). FBI IC3 data show cybercrime losses exceeding $10.3 billion in 2022, with Business Email Compromise as the single largest category; IBM research places the average cost of a social-engineering-initiated breach above $4 million (FBI IC3, 2023; Gururaj et al., 2024). Existing reviews predate the operational deployment of generative AI and deepfake synthesis as attack tools, and no integrated framework currently bridges psychological theory, an updated attack taxonomy, and AI-specific defense design. Drawing on Hadnagy (2011), Gray (2021), Sonowal (2023), and Gururaj et al. (2024), this paper traces social engineering from its historical origins through its AI-augmented present, maps the current attack taxonomy, examines the psychological mechanisms that make these attacks effective, and proposes a defense-in-depth framework that treats human behavior and technical controls as co-equal requirements. The paper's primary contribution is a synthesis connecting behavioral science, organizational design, and emerging AI threats in a single analytical frame applicable to practitioners and policymakers.

Keywords. Social engineering, phishing, deepfake, AI-augmented attacks, defense-in-depth.

1. Introduction

Ask a security engineer where breaches come from, and the answer will usually involve unpatched systems, misconfigured firewalls, or compromised credentials. Ask the attackers, and the answer is simpler: people. Social engineering, the art of manipulating individuals into surrendering access, information, or authorization they would otherwise protect, has been the dominant breach vector for over a decade, and its share of the attack surface is growing rather than shrinking. As technical defenses have matured, the return on investment from psychological manipulation has outpaced the return from technical exploitation. Hadnagy (2011, p. 8) states the logic plainly: no attacker is going to spend a hundred hours on a technical route when a well-crafted phone call achieves the same result in under an hour.

The FBI's IC3 reported losses exceeding $10.3 billion from cybercrime in 2022, with Business Email Compromise accounting for the single largest category of financial harm (FBI IC3, 2023). An estimated 98% of cyberattacks rely on social engineering at some point in the kill chain (Sonowal, 2023, p. 2). IBM's 2022 breach cost data found that social engineering as an initial attack vector produced average incident costs above $4 million, exceeding the mean for purely technical exploits (Gururaj et al., 2024). What transforms these figures from alarming to urgent is the trajectory: the convergence of social engineering with generative AI, voice and video synthesis, and automated open-source intelligence collection is producing a

threat landscape that the defense frameworks built over the past twenty years were not designed to handle.

No existing single-paper synthesis bridges psychological theory, a taxonomy updated to include AI-augmented vectors, and a defense framework designed for the current threat environment. This paper addresses that gap. The analysis draws on Hadnagy (2011), Gray (2021), Sonowal (2023), and Gururaj et al. (2024) as its primary sources, supplemented by FBI IC3 data, IBM breach cost research, and documented incident records.

2. Methodology

This paper employs a systematic narrative review methodology. Source selection followed three criteria: theoretical centrality, empirical grounding, and recency, with preference given to sources published after 2020 for empirical claims about attack prevalence and cost. Four primary texts form the theoretical core: Hadnagy (2011), which provides the foundational taxonomy and psychological framework; Gray (2021), which addresses practitioner methodology and OSINT-based attack construction; Sonowal (2023), which synthesizes the historical and definitional landscape alongside contemporary threat data; and Gururaj et al. (2024), which integrates AI-augmented attack analysis with organizational defense recommendations.

Empirical data on incident frequency, financial losses, and breach costs are drawn from the FBI IC3's annual Internet Crime Reports, IBM Security's Cost of a Data Breach series, and the CSIS Significant Cyber Incidents database. Three case studies were selected to maximize attack vector diversity: the 2011 RSA SecurID breach, the 2020 Twitter Bitcoin scam, and the 2024 Hong Kong deepfake fraud. This paper does not present primary empirical data; it is a synthesis and framework paper.

3. Historical and Conceptual Background

The history of social engineering is considerably older than the internet. The psychological mechanisms attackers exploit today are stable features of human cognition that have been exploited across every era of recorded history; they cannot be patched or engineered away. The Trojan Horse is the oldest documented case study: the Greeks did not breach Troy's walls, they persuaded the Trojans to breach them on the Greeks' behalf (Sonowal, 2023, p. 3). The term itself entered the English lexicon from Dutch industrialist J.C. Van Marken, who used sociale ingenieurs in 1894 to describe specialists who applied systematic discipline to the human dimensions of organizations (Sonowal, 2023, p. 3).

The twentieth century produced practitioners who advanced the discipline substantially. Victor Lustig's 1925 Eiffel Tower scheme combined forged documentation, a choreographed credibility-building process, and manufactured urgency in a sequence that modern BEC operators replicate almost exactly (Sonowal, 2023). Kevin Mitnick's penetrations of Nokia, Motorola, and U.S. Defense Department networks established that psychological acuity combined with technical knowledge produced an attack surface no firewall could address. As Hadnagy (2011, p. 8) observes, the economics have only become more favorable to attackers as technical defenses have hardened.

This paper uses Hadnagy's (2011) definition of social engineering: any act that influences a person to take an action that may or may not be in their best interest. Sonowal (2023) draws the important distinctions: social engineering differs from pure technical hacking in that it targets people rather than systems; from simple deception in that it involves a structured, multi-phase operational sequence. The distinction between social engineering and

phishing deserves particular emphasis, because treating them as synonymous produces defenses that address only a fraction of the actual attack surface.

4. The Psychological Architecture of Manipulation

Social engineering works reliably not because attackers are especially skilled manipulators but because they have learned to activate cognitive mechanisms that human beings share universally. Understanding the psychological architecture underlying social engineering is the prerequisite for designing defenses that actually interrupt attacks rather than simply making them marginally harder.

4.1 Cialdini's Principles of Influence

Robert Cialdini's (2006) six principles of influence remain the most cited and most operationally relevant framework in the social engineering literature. Gray (2021), Hadnagy (2011), and Sonowal (2023) all draw on them as primary analytical tools. Reciprocity creates a felt obligation to return favors; attackers activate it by providing something of apparent value before making a request. Authority triggers compliance with perceived power figures.

Social proof leads people to take behavioral cues from what they believe others are doing. Scarcity and urgency engage loss aversion directly. Liking raises compliance with people perceived as similar or warm. Commitment and consistency mean that a small initial agreement makes subsequent, larger compliance significantly more likely (Cialdini, 2006; Hadnagy, 2011).

4.2 Cognitive Biases

Alongside Cialdini's principles, attackers exploit a specific cluster of cognitive biases. The optimism bias leads individuals to underestimate their own probability of being targeted. The automation bias produces over-reliance on technical filters. The familiarity heuristic means that a sender name or visual identity that approximates a trusted source triggers approval before deeper scrutiny engages (Gururaj et al., 2024). Most consequentially, Hadnagy (2011) documents the deception detection deficit: human beings perform at approximately chance level — around 54% accuracy — when asked to identify deceptive communications. Any defense strategy that relies on trained employees detecting malicious intent from communication content alone is built on a foundation the research does not support.

4.3 Elicitation

Elicitation is the subtle extraction of information during an apparently normal and innocent conversation (Hadnagy, 2011, p. 98). Where phishing requires the target to perform a conscious action, elicitation produces voluntary disclosure: the target provides sensitive information without recognizing they are doing so. Hadnagy (2011, p. 99) documents a field example in which a complete operational picture of a target organization — including staffing structure, system names, and security practices — was assembled at a professional networking event through what the target believed was routine conversation.

5. A Comprehensive Attack Taxonomy

Social engineering is not a single technique but a structured family of tactics sharing psychological foundations while varying substantially in medium, technical requirements, and scale. The taxonomy below synthesizes classifications from Sonowal (2023), Gray (2021), Gururaj et al. (2024), and Hadnagy (2011), updated to reflect vectors that have reached operational maturity since those texts were written.

5.1 The Attack Lifecycle

Sonowal (2023) describes a six-phase operational lifecycle: (1) target selection based on role and access level; (2) OSINT-based information gathering using social media analysis, public records, Google dorking, and EXIF metadata extraction; (3) preparation of the pretext and delivery infrastructure; (4) trust development, which can require considerable patience (Gururaj et al., 2024, p. 33); (5) exploitation; and (6) exit and cover. By the time an attack reaches the exploitation phase, the psychological preparation is already complete. Effective defense must reach back into the reconnaissance and trust-development phases, not merely the moment of the lure.

5.2 Digital Attack Vectors

Phishing remains the highest-volume social engineering vector. Mass phishing distributes deceptive emails at scale; spear phishing trades volume for precision by incorporating OSINT-derived specifics. Gray (2021) describes constructing pretexts around live events in the target's environment. Whaling applies spear phishing precision to C-suite executives. Vishing adds a real-time interactive dimension that Hadnagy (2011) argues makes it more effective against security-trained recipients. Smishing exploits higher SMS open rates. Business Email Compromise impersonates internal executives to authorize wire transfers; the FBI IC3 reported $2.7 billion in adjusted BEC losses in 2022 alone (FBI IC3, 2023).

Pretexting is foundational to virtually all social engineering. Hadnagy (2011) makes the case that every attack involves a fabricated scenario providing the attacker's request with plausible operational context. Physical social engineering- tailgating, dumpster diving, baiting with infected USB drives- is consistently underweighted in digital-first security frameworks, despite consistently succeeding in authorized penetration tests (Gray, 2021; Gururaj et al., 2024).

5.3 Target Demographics

Social engineering attacks are not randomly distributed. Sonowal (2023) shows that systematic profiling by role, access level, and information value is standard practice. Entry-level employees are targeted for their lack of institutional skepticism; senior executives for their authorization authority over wire transfers and policy overrides; help desk and IT staff specifically because their function involves granting access and resetting credentials.

6. The AI-Augmented Threat Landscape

Artificial intelligence has not altered the psychological principles that social engineering exploits. What it has changed is the cost, speed, and scale at which sophisticated attacks can be executed. Gururaj et al. (2024) characterize this as a qualitative shift rather than a quantitative increment.

6.1 AI-Generated Phishing

Large language models have effectively eliminated the surface-level anomalies that security awareness training has relied on for detection: grammatical errors, awkward phrasing, implausible sender identities. A model prompted with a target's LinkedIn profile and publicly available corporate communications can produce a spear-phishing message referencing authentic organizational events, real colleague names, and genuine internal systems.

Gururaj et al. (2024) describe the consequence: the gap between mass phishing and spear

phishing is collapsing. AI now makes it operationally viable to run campaigns that have both volume and precision.

6.2 Deepfake Audio and Video

In 2019, a UK energy company transferred approximately £200,000 after the CEO received a call from a synthesized voice replicating his parent company's Managing Director. In February 2024, a Hong Kong-based multinational lost approximately HK$200 million (USD $25.6 million) when a finance employee authorized transfers following a video conference in which every other participant was an AI-generated deepfake constructed from publicly available footage of real executives. Hong Kong police confirmed this as the first publicly documented case of multi-participant deepfake personas used simultaneously in a real-time interactive fraud (Gururaj et al., 2024).

6.3 Automated OSINT

AI has compressed OSINT reconnaissance from hours to seconds. Automated tools now cross-reference social media profiles, corporate directories, breach databases, and news archives to produce detailed targeting dossiers at scale. Gray (2021) demonstrates that photograph EXIF metadata alone can yield GPS coordinates, device model, and operating system information. The normalization of remote work and proliferation of IoT devices have further expanded the attack surface: home networks lack enterprise-grade segmentation, and IoT devices run on weak default credentials (Gururaj et al., 2024).

7. Case Studies

7.1 The 2011 RSA SecurID Breach

In March 2011, RSA Security was breached through a single phishing email with the subject line "2011 Recruitment Plan.xls." The attachment exploited a then-unpatched Adobe Flash zero-day. Three psychological mechanisms were engaged simultaneously: relevance (recruitment information was professionally salient), credibility (the apparent sender was plausible), and curiosity (a salary spreadsheet implied actionable information). The Remote Access Trojan installed through that single interaction gave attackers access to SecurID seed values subsequently used to compromise Lockheed Martin and other defense contractors. One employee's momentary decision cascaded into a national security incident (Hadnagy, 2011; Sonowal, 2023).

7.2 The 2020 Twitter Bitcoin Scam

In July 2020, attackers compromised Twitter's internal administrative tools without exploiting a single technical vulnerability. Access was obtained entirely through a coordinated vishing campaign: the attackers impersonated internal IT staff, directed targets to a fake VPN portal, and collected credentials through a process that looked operationally normal to the employees. With administrative access, they hijacked verified accounts of Barack Obama, Joe Biden, Elon Musk, Bill Gates, and Apple, running a Bitcoin scam returning approximately $120,000 in under thirty minutes (Sonowal, 2023). The case is important precisely because Twitter's technical security was adequate; the attack succeeded because technical security cannot protect against threats operating through the authorized humans who administer it.

7.3 The $25 Million Hong Kong Deepfake Fraud (2024)

The February 2024 Hong Kong case represents a step-change in financial scale and technological sophistication. A finance employee authorized transfers totaling approximately HK$200 million after participating in a video conference in which every other participant was an AI-generated deepfake of real senior executives. The employee became suspicious only after contacting the real individuals through a separate channel, by which point all transfers had been executed. Hong Kong police confirmed this as the first publicly documented case of multiple deepfake personas deployed simultaneously in a real-time interactive fraud (Gururaj et al., 2024).

8. An Integrated Defense Framework

Social engineering cannot be addressed by any single countermeasure. The attacks operate across technical, psychological, procedural, and physical dimensions simultaneously. The framework proposed below synthesizes recommendations from Gray (2021), Sonowal (2023), Hadnagy (2011), and Gururaj et al. (2024) into three interlocking components.

8.1 Technical Controls

Email authentication infrastructure offers the highest technical leverage against phishing. Gray (2021) presents the SPF, DKIM, and DMARC protocol stack as the foundational layer; organizations deploying DMARC at enforcement level (p=reject) effectively eliminate domain spoofing as a viable attack vector. Hardware security keys implementing FIDO2/WebAuthn provide the strongest MFA protection because they are physically bound to the authentication session and cannot be relayed by adversary-in-the-middle attacks. Gururaj et al. (2024) recommend AI-powered behavioral analytics capable of flagging anomalous communication patterns and synthetic audio signatures. Zero Trust architecture requires continuous verification of every access request regardless of origin.

8.2 Human-Centered Defense

Gray (2021) draws a sharp distinction between compliance-oriented training — the annual awareness module satisfying a regulatory checkbox — and engagement-oriented programs that develop actual threat recognition capability. Programs producing measurable outcomes: run continuously rather than annually; maintain a non-punitive culture (Gray, 2021, p. 21, explicitly argues that public identification of employees who fail simulated attacks suppresses incident reporting and destroys the psychological safety effective defense requires); use behavioral simulation rather than passive knowledge transfer; differentiate content by role; and update continuously to address AI-generated lures.

The core behavioral defense that Hadnagy (2011), Gray (2021), and Sonowal (2023) converge on is verify out-of-band. Any request for a sensitive action — a wire transfer, a credential reset, a grant of system access — must be confirmed through a separate, pre-established communication channel before it is executed. This single practice, applied consistently, neutralizes even the most technically sophisticated attacks. The 2024 Hong Kong deepfake fraud would have been stopped at this step.

8.3 Organizational Governance and Culture

Sonowal (2023) identifies three primary governance domains requiring explicit policy: internet usage, email security, and physical access. Critical elements include authorization procedures for sensitive transactions that cannot be bypassed regardless of the apparent seniority of the requester; least-privilege access controls; and clear, non-punitive escalation procedures. Gururaj et al. (2024) identify unclear reporting structures as a significant

organizational vulnerability: an employee who suspects manipulation but fears blame will delay or avoid reporting, collapsing the response window.

All three pillars rest on organizational culture. Gray (2021), Hadnagy (2011), and Gururaj et al. (2024) reach consistent conclusions: the strongest single predictor of social engineering resilience is whether employees genuinely feel safe to report suspicious contacts and challenge apparent authority. Edmondson's (1999) research on psychological safety provides the theoretical support: organizations with high psychological safety detect errors faster and develop more durable competence over time.

9. Discussion

The most consequential strategic implication of AI-augmented social engineering is that the dominant model of security awareness training is structurally obsolete. Programs built around teaching employees to recognize surface-level anomalies were designed for a threat environment that no longer exists. A large language model prompted with a target's public professional profile can produce a spear-phishing message that clears every heuristic a trained employee has been taught to apply. Training people to detect AI-generated content is not a viable alternative; the necessary reorientation is from message-level detection to process-level verification.

The culture finding carries its own strategic implication. Edmondson's (1999) psychological safety research shows that organizations where people feel safe to admit mistakes and report concerns detect incidents earlier and recover more effectively. Organizations that treat every failed phishing simulation as a performance failure produce a workforce that conceals incidents rather than reporting them. Measurement frameworks for security awareness programs need to track reporting rates alongside failure rates. Legal and regulatory frameworks have not kept pace with the threat; Sonowal (2023) finds that the psychological manipulation component of social engineering frequently falls into legal grey areas until material harm has already occurred.

10. Limitations

As a systematic narrative review rather than an empirical study, this paper's conclusions depend on the quality and representativeness of the source material. The four primary texts do not represent the full breadth of the social engineering literature, particularly work published in non-English languages. The case studies are limited to three publicly documented incidents; the full population of social engineering attacks is substantially larger and includes many incidents that are never publicly disclosed. Empirical data are drawn from self-reported incident databases and vendor-sponsored research, both of which have well-documented biases. The AI-threat analysis is based on publicly disclosed incidents as of early 2025; specific capability threshold claims may require updating.

11. Future Research Directions

Longitudinal empirical studies comparing compliance-oriented and engagement-oriented security awareness training against documented incident rates are needed to move the field from practitioner intuition to evidence-based program design. The psychological safety and security culture relationship identified by Edmondson (1999) has not been tested empirically in organizational cybersecurity contexts. The AI threat landscape requires dedicated empirical work: the 2024 Hong Kong deepfake fraud is currently the most thoroughly documented multi-participant synthetic media attack, but systematic incident

analysis across a larger population of AI-augmented cases does not yet exist. A comparative legal analysis of how existing criminal and civil frameworks apply to the manipulation techniques used in social engineering would provide a foundation for the regulatory evolution the field clearly needs (Sonowal, 2023).

12. Conclusions

Social engineering is the oldest attack method on record, and artificial intelligence has made it more effective than it has ever been. Its persistence across every era comes down to a structural reality: every technical system is ultimately operated by human beings whose cognition is built around trust, cooperation, and responsiveness to authority. Those tendencies are not vulnerabilities in the engineering sense; they are the foundations of functional social life. The question is not whether they can be removed but whether organizations can build systems and cultures that make them harder to exploit systematically.

This paper has made three substantive contributions. First, it identified a gap in the existing literature: no prior synthesis addresses the intersection of behavioral psychology, an updated attack taxonomy including AI-augmented vectors, and a defense framework designed for the current threat environment. Second, it proposed a defense-in-depth framework that treats technical controls, human-centered training, and organizational governance as co-equal components, with organizational culture as the foundation that determines whether the other components function as intended. Third, it traced the AI escalation through documented incidents to establish that the shift from theoretical threat to operational reality has already occurred. As Hadnagy (2011, p. 8) put it, the bad guys do not stop because of a contractual limitation or their own morals. The only proportionate response is a commitment to understanding, learning, and building human resilience that is equally persistent and equally unconstrained.

References

1. Cialdini, R. B. (2006). Influence: The Psychology of Persuasion (revised ed.). Harper Business.
2. Edmondson, A. C. (1999). Psychological safety and learning behavior in work teams. Administrative Science Quarterly, 44(2), 350–383. https://doi.org/10.2307/2666999
3. Federal Bureau of Investigation — Internet Crime Complaint Center (IC3). (2023). 2022 Internet Crime Report. FBI. https://www.ic3.gov/Media/PDF/AnnualReport/2022_IC3Report.pdf
4. Gray, J. (2021). Practical Social Engineering: A Primer for the Ethical Hacker. No Starch Press.
5. Gururaj, H. L., Janhavi, V., & Ambika, V. (Eds.). (2024). Social Engineering in Cybersecurity: Threats and Defenses. CRC Press. https://doi.org/10.1201/9781003406716
6. Hadnagy, C. (2011). Social Engineering: The Art of Human Hacking. Wiley.
7. IBM Security. (2022). Cost of a Data Breach Report 2022. IBM Corporation. https://www.ibm.com/reports/data-breach
8. Mitnick, K. D., & Simon, W. L. (2002). The Art of Deception: Controlling the Human Element of Security. Wiley.
9. Mouton, F., Leenen, L., & Venter, H. S. (2016). Social engineering attack examples, templates and scenarios. Computers & Security, 59, 186–209. https://doi.org/10.1016/j.cose.2016.03.004
10. Purplesec. (2021). 2021 Cyber Security Statistics: The Ultimate List of Stats, Data & Trends. Purplesec LLC. https://purplesec.us/resources/cyber-security-statistics/
11. Sonowal, G. (2023). Social Engineering Attack: Rethinking Responsibilities and Solutions. Nova Science Publishers. https://doi.org/10.52305/KSOA7898
12. Twitter, Inc. (2020, July 30). An update on our security incident. Twitter Safety Blog. https://blog.twitter.com/en_us/topics/company/2020/an-update-on-our-security-incident
13. Van Marken, J. C. (1894). Sociale ingenieurs. As cited in Sonowal, G. (2023). Nova Science Publ.

Why We Need New Access Models: From Session-Bound Authorization to Continuous Access

Ghenadie Belinschi

Academy of Economic Studies of Moldova, Chisinau, Moldova
ghenadie.belinschi@ase.md

Abstract. This article shows why existing security approaches and access-control models no longer meet the requirements of today's digital environment, and what next-generation models should incorporate to close the resulting gap. We briefly review classical models (DAC/MAC, RBAC, ABAC, etc.) and identify where they fall short amid highly dynamic context, cross-organizational workflows, and compressed decision-making cycles. On this basis, we articulate requirements for next-generation models: continuous, context-adaptive authorization; explicit treatment of risk and trust; machine-readable and verifiable policies; and traceable delegation with revocation. Practical examples illustrate the applicability of several models, along with their limitations and risks (e.g., staleness of contextual attributes, increased administrative complexity, reduced explainability). Our central thesis is a shift from replacement to composition: classical mechanisms retain their value as a stable foundation, while new prototypes are layered on top to close gaps without sacrificing manageability. The article offers practical guidance for designing and evolving access-control systems and security policies in modern organizations.

Keywords. Access control; dynamic context; risk-adaptive authorization; model composition; security policy.

1. Introduction

Over the past half century of access control, the IT industry has moved from simple permission matrices to mature enterprise systems based on roles and attributes. These solutions shaped the security vocabulary (subject, object, operation), delivered manageability and regulatory compliance, and enabled policy portability across systems [1], [6], [7], [8]. However, computing architectures have changed radically: data resides in clouds and on edge devices; organizations operate across domain and jurisdictional boundaries; employees move between networks and applications; and the "last mile"—the moment when information is rendered to a human—has become a vulnerability in its own right.

In this setting, traditional session-bound authorization ("check once—allow until the session expires") is no longer adequate. A new class of models is needed in which the access decision is not fixed but is maintained and revised throughout the entire interaction with the data; in which the object itself remains encrypted until policy conditions are satisfied; and in which relationships among participants are as important as their statuses [9], [10], [12]–[14]. The goal of this paper is to show why existing approaches no longer meet modern requirements and what new models can introduce to close the gap. We draw on established paradigms (DAC/MAC, RBAC, ABAC/UCON, Zero Trust, ReBAC) and compare them with the challenges of distributed infrastructures, interorganizational cooperation, and last-mile exposure [1], [4]–[12].

2. Historical Foundations

Early Foundations: Protection Matrix and Capabilities

The first formalizations described the world simply: subjects and objects whose relationships are recorded in a matrix, and a process that carries with it a set of capabilities—roughly, tokens marking the right to perform a given operation [3]. This era gave us the language and mechanics we still use today: subject/object/operation, the representation of rights as data, and the idea of delegation. But as systems grew, it became clear that a flat matrix balloons too quickly, and managing delegations becomes no less complex than the business processes themselves [1], [3].

Two Streams of the 1970s–1990s: Discretionary and Mandatory

The mass-market IT answer to the question "who is the owner?" was Discretionary Access Control (DAC): the object owner decides who can do what; thus the familiar ACLs, groups, and the UNIX/Windows "file owner" emerged. This is flexible and administrator-friendly but vulnerable to Trojan-horse attacks and hard to oversee at scale [1]. In parallel, the military-industrial domain insisted on a different logic—Mandatory Access Control (MAC): secrecy labels and formally provable information-flow constraints (Bell–LaPadula lattices for confidentiality and Biba for integrity) [4], [5]. MAC makes leakage "mathematically hard," but fits poorly with open collaboration and the everyday fluidity of access.

1990s–2000s: Manageability through Roles

The corporate IT revolution created the need to manage functions rather than individual permissions. RBAC took shape accordingly: roles aggregate permissions; users are assigned roles; and separation of duties, role administration, and an ANSI standard followed [6]. RBAC delivered manageability and audit compliance, but encountered a "role explosion" across diverse business cases, where hundreds of nuances turned the role table into a new matrix.

2000s–2010s: Expressiveness through Attributes and Usage Control

The focus shifted from "who you are" to "what is known about you and what the context is." ABAC (and the XACML ecosystem) allowed policies to be expressed as predicates over subject, object, and environment attributes, while separating the Policy Decision Point (PDP) from Policy Enforcement Points (PEP) [7], [8]. This increased expressiveness but raised questions about attribute provenance and quality, as well as testability and explainability. A further step was UCON: access as use under a set of obligations and conditions, with attributes changing on the fly [9]. Access ceased to be a "flick of a switch" and became a process dependent on the fulfillment of obligations and the dynamics of context.

2010s–2020s: Zero Trust, Relationship Graphs, and the Vanishing Perimeter

Mobility, SaaS, and microservices eroded the notion of a "trusted inside." Zero Trust Architecture prescribed revalidating every request, microsegmenting access, and treating the network as hostile by default [10]. In collaborative products, another regularity emerged: permissions are naturally derived from relationships—"workspace member," "folder editor," "delegated by X." This led to system-scale implementations of ReBAC (relationship-based access control), where a global authorization graph becomes the primary source of truth [12]. The efficiency gains are dramatic: instead of hundreds of role combinations, a small set of edge types and path checks suffices.

3. Brief Context: What Classical Models Solved

Discretionary and mandatory access control were developed under different design logics: DAC grants the object owner the authority to decide "who can do what," providing convenience and flexibility; MAC applies secrecy/integrity labels and enforces formally provable information-flow restrictions, which critical systems require [1], [4], [5]. The corporate era introduced RBAC—roles as a "container" of permissions and manageability at enterprise scale [6]. As access scenarios grew more complex, attribute-based policies (ABAC) and the separation of Policy Decision Points (PDP) from Policy Enforcement Points (PEP) emerged—greater expressiveness and centralized policy control with distributed enforcement [7], [8]. In parallel, the idea of Usage Control (UCON) evolved: authorization is not a one-time admission check but a process tied to obligations and conditions that may change on the fly [9]. As the network "perimeter" ceased to be a reliable boundary, the Zero Trust paradigm appeared: trust neither networks nor hosts by default; verify each request [10]. In highly collaborative systems, relationships moved to the forefront—who participates in which space, who delegated to whom—and permissions are derived from a relationship graph (ReBAC) [12]. Each paradigm addressed the problems of its time and remains valuable within its scope. Today's environment, however, imposes three classes of requirements that exceed those boundaries.

4. Where Current Approaches Fall Short

First, contextual dynamics. The "static session" has given way to a stream of events: a device switches networks, behavioral risk rises, biometric "liveness" drops out. Permission granted in the morning may be unsafe by midday. Session-bound authorization cannot loosen its grip as conditions change, and classical extensions (hidden timeouts, idle thresholds) are merely surrogates for true contextual dynamics [9], [10]. Second, interorganizational relationships. Projects now bring together people and services from different legal entities, domains, and cloud tenants. Role models designed for a single organization lead to role explosion and exceptions, while attribute-based models complicate the issuance and verification of attributes for external participants. In practice, rights are often determined by relationships (participation, ownership, delegation), which are naturally expressed as a graph rather than flat lists [12]. Third, perimeter dissolution and last-mile vulnerability. Data are copied, cached, and flow to endpoints and third-party clouds. Even perfect server-side authorization is insufficient if an object appears in plaintext on the wrong screen: shoulder-surfing, a compromised OS, or screen recording can all slip past policy. Traditional access models do not even perceive the moment of display and cannot manage it as part of access control.

5. What New Models Must Provide

1). Continuity and risk adaptiveness. Access is not an entry pass but a state maintained as long as context remains within the risk policy. The authorization decision can be revised at any moment: if verified user presence ceases, access collapses; if the connection shifts to an unsafe network, permissions automatically narrow. This requires embedding—mechanically and semantically—the notions of revocation and controlled "weakening" of access in response to risk signals [9], [10].

2). "Data-as-controller". An object must be able to withhold disclosure unless conditions are met. This is achieved by moving part of the control into the data: attribute-/factor-dependent encryption (including biometrics) and policies defined at the object level.

Until the predicate holds, it remains ciphertext; once satisfied, decryption occurs locally for the minimum necessary time [13], [14].

3). Relationship semantics. Actual permissions are determined by who is related to whom and how. The model must derive permissions from a relationship graph (space member, co-owner, delegate) and validate them with guaranteed consistency at scale [12]. This reduces administration and reflects collaboration better than role combinatorics.

4). Display control as part of access. The last mile cannot remain outside policy: where and how plaintext is rendered, to whom, and through which channel. One-way, personal output channels (e.g., a standalone wearable without network interfaces that locally decodes ciphertext) are critical, as is binding verified "live presence" to the very permission to display [10], [15].

5). Revocability and replaceability of bindings. Compromising auxiliary data should not permanently compromise the biometric or require changing the object - reissuing the binding suffices. The cryptographic layer must be replaceable (crypto-agile): policy is independent of specific primitives and can withstand transition to post-quantum schemes [13], [14].

6. Illustrations of the Requirements in Typical Scenarios

Distributed analytics scenario. A team from three companies prepares a joint report. Data are replicated across clouds, and participants come and go. Role-based models within individual organizations produce a "combinatorial explosion" of roles and delegations; attribute-based models provoke disputes over who issued attributes and how they expire. A relationship graph expresses participation and delegation; "data-as-controller" ensures that a downloaded report remains ciphertext outside the project context; continuity enables automatic removal of rights as soon as a participant leaves the project [7], [10], [12], [13].

Mobile access to sensitive data scenario. An employee reads a document at an airport. One-time authorization at 9:00 a.m. does not account for the fact that at 9:05 a.m. he joined an unfamiliar open network, and at 9:07 a.m. people gathered around him. Continuous control lowers the access level when the network changes; "last-mile" protection blocks display without a personal decoder device; the object remains ciphertext on a shared PC screen [9], [10].

"Time-bound delegation" scenario. A manager goes on vacation and temporarily delegates access to a folder with contracts to an assistant. Role-based assignment of permissions risks being forgotten later. The relationship graph encodes delegation with an automatically expiring edge; "data-as-controller" will not reveal itself to the assistant unless contextual conditions (location, device) are met [12], [13].

In each example, the object is encrypted under stated conditions, relationships are explicitly expressed and verified, and the access decision is not one-off but sustained.

7. Manageability and Explainability

New models are often criticized as "black boxes." That critique is appropriate when policies are scattered across systems. Properly organized, however, they improve explainability: policy is expressed as code (policy-as-code); a single source of truth is versioned and tested; the relationship graph provides verifiable paths; and cryptographic binding yields an unambiguous condition for why decryption occurred. For audit, this is even more convenient than disparate ACLs and exotic roles [7], [11], [12], [13].

It is also important that the new principles are compatible with existing ones: RBAC remains a convenient "add-on" as manageability layer, ABAC provides expressive condition

logic, and Zero Trust enforces network-level discipline. This is not a replacement but a composition.

8. Limitations and Risks to Keep in Mind

First, the hygiene of attributes and relationships is critical: who issues them, how long they remain valid, and how they are revoked. Without disciplined provenance, attribute-based policies become a source of errors [7].

Second, privacy. Any solution that uses biometric factors must guarantee the revocability of bindings and the non-reconstructability of biometrics from auxiliary data; fuzzy extractors address exactly this requirement [14].

Third, post-quantum resilience. If control is pushed into the cryptographic layer, the system must allow primitives to be replaced without rewriting policies. This argues for clear layering: policy separate from cryptography [13].

Finally, "last mile engineering is required". Personal display devices should be widely available and manageable. The market for wearable glasses and AR devices is evolving rapidly, which makes this aspect practically achievable [2].

9. What Is Relevant Today

Some of the ideas described are already being implemented in products: global authorization services built on relationship graphs (inspired by industrial-grade systems like Zanzibar), centralized attribute-based policy engines, and "Zero Trust" as a discipline of re-verifying each request [10], [12]. Policy-as-code engineering practices are taking root in secure development, enabling policies to be tested and explained [7], [11]. Cryptographic "data-as-controller" approaches are being actively researched and deployed in niche scenarios (attribute-based encryption, confidential containers) [13]. Against this backdrop, a natural next step is to bind access to "live" factors - including biometric ones - with guaranteed revocability, and to govern the visualization itself as part of policy [14], [15], [2].

10. Conclusion

Traditional access-control models brought order in an era of centralized systems and stable roles. But where data are distributed, participants are linked by flexible relationships, and the threat shifts to the moment of display, session-bound ("one-time") authorization stops working. New models must do three things: (1) turn access from an event into a continuous state with revocation driven by risk signals; (2) move part of the control into the object so that it does not disclose outside the required conditions; and (3) treat relationships as the primary source of meaning for authority. These principles do not replace RBAC and ABAC; they compose with them into a coherent architecture.

This perspective opens the way to practices in which the human, their relationships, and the data themselves form a single security framework. On this foundation, further development is possible—even solutions in which an encrypted document "lives" inside a biometrically bound carrier and is displayed only through a personal, one-way channel. New access models are needed not "for fashion's sake," but because they are the only ones capable of faithfully accounting for contextual dynamics, perimeter dissolution, and last-mile vulnerability - without sacrificing manageability and explainability [1], [7], [10], [12]–[14].

References

1. Benantar, M. Access Control Systems: Security, Identity Management and Trust Models. Springer, 2006. 356 p. ISBN 978-0-387-26166-5.
2. "Driving Needs for New Methods; The Market for Wearable Offline Devices for Secure Display." Label: CRD-NEW-MODELS-V1. Author's archive.
3. Lampson, B. W. Protection. ACM Operating Systems Review 8(1), 1974, 18–24.
4. Bell, D. E.; LaPadula, L. J. Secure Computer Systems: Mathematical Foundations. MITRE MTR-2547, 1973–1975.
5. Biba, K. J. Integrity Considerations for Secure Computer Systems. MITRE MTR-3153, 1977.
6. Sandhu, R.; Coyne, E.; Feinstein, H.; Youman, C. Role-Based Access Control Models. IEEE Computer 29(2), 1996, 38–47; ANSI/INCITS 359-2004.
7. Hu, V. C.; Ferraiolo, D.; Kuhn, R. Guide to Attribute-Based Access Control (ABAC). NIST SP 800-162, 2014.
8. OASIS. eXtensible Access Control Markup Language (XACML) Version 3.0. OASIS Standard, 2013.
9. Park, J.; Sandhu, R. The UCON ABC Usage Control Model. ACM Transactions on Information and System Security (TISSEC) 7(1), 2004, 128–174.
10. Rose, S.; Borchert, O.; Mitchell, S.; Connelly, S. Zero Trust Architecture. NIST SP 800-207, 2020.
11. NISTIR. Practices for describing and verifying access-control policies (policy-as-code) in secure development.
12.] Google. Zanzibar: Google's Consistent, Global Authorization System. USENIX ATC, 2019 (ReBAC/graph-based authorization).
13. Bethencourt, J.; Sahai, A.; Waters, B. Ciphertext-Policy Attribute-Based Encryption. IEEE Symposium on Security and Privacy, 2007.
14. Dodis, Y.; Reyzin, L.; Smith, A. Fuzzy Extractors: Strong Keys from Biometrics. SIAM Journal on Computing 38(1), 2008, 97–139.
15. Atkinson, G. A.; Smith, L. N. Using Photometric Stereo for Face Recognition. Image and Vision Computing 28(5), 2010, 806–816.

Global Trends in Intellectual Property in 2023–2024

Grisciuc-Bucica Svetlana

Academy of Economic Studies of Moldova, Chisinau, Moldova
svetlana_lawyer@yahoo.com

Abstract. This article examines global trends in the field of intellectual property (IP) based on statistical data published by the World Intellectual Property Organization (WIPO), including the reports IP Facts and Figures 2024-2025 and World Intellectual Property Indicators 2024-2025. The study analyzes application dynamics across patents, utility models, trademarks, and industrial designs, focusing on developments in 2023–2024. Particular attention is paid to regional distribution patterns, application intensity relative to GDP, sectoral specialization, and the increasing role of middle-income countries in the global IP system. The analysis reveals continued growth in intellectual property activity, particularly in Asian economies, as well as a gradual increase in the share of resident filings in developing jurisdictions. The findings also indicate a structural shift toward technology- and design-intensive sectors and an expanding role of intellectual property indicators in assessing innovation performance. By combining statistical analysis with a policy-oriented perspective, the article provides an updated view of global intellectual property dynamics and discusses their implications for the development of national IP strategies. Particular attention is given to the role of institutional incentives, sectoral priorities, and the integration of IP statistics into economic and innovation policymaking.

Keywords. Intellectual property, WIPO, patents, trademarks, utility models, industrial designs, IP statistics, innovation policy, GDP, middle-income countries

1. Introduction

The contemporary knowledge-based economy predetermines intellectual property (IP) rights as one of the key resources for the technological and social development of modern society. Having acquired the status of a strategic asset, intellectual property is increasingly moving beyond the boundaries of the purely legal sphere and is becoming a central element of the innovation, industrial, and digital policies of modern states.

The analysis of global statistical data published by the World Intellectual Property Organization (WIPO)[1], including the reports IP and Facts 2024[2], IP Facts and Figures 2025 [3], World Intellectual Property Indicators 2024[4] and World Intellectual Property Indicators 2025[5], provides a unique opportunity to identify the key directions in the transformation of the global system for the protection of intellectual property rights. The focus of this study is not limited to quantitative changes in the main categories of industrial property rights, but also includes their regional distribution, the structure of sectoral demand, as well as the intensity of filing activity in relation to the economic potential of different countries.

Despite the availability of extensive international statistics on intellectual property, a systematic interpretation of recent global filing trends and their policy implications remains relatively limited in the academic literature. This article seeks to contribute to this discussion by providing an integrated analysis of key statistical indicators and emerging structural shifts in global intellectual property activity.

2. Global Trends in Intellectual Property Activity (2023–2024)

A detailed analysis of global filing activity in 2023 indicates that the international system of industrial property demonstrated positive dynamics across most key categories of intellectual property rights. According to WIPO data[1], the number of patent applications increased by 2.7%, utility model applications by 3.9%, and industrial design applications by 2.8%. At the same time, the number of trademark applications declined by approximately 2% compared with the previous year. Nevertheless, the overall volume of filings remained consistently high, reaching around 15.2 million trademark applications, 3.6 million patent applications, and approximately 1.5 million industrial design applications worldwide.

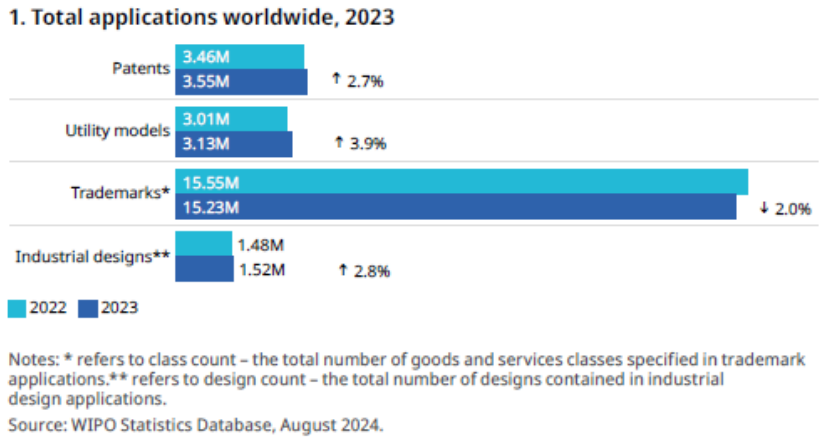


Fig. 1. Total applications worldwide, 2023.

Data for 2024 confirm the continuation of these trends. According to the WIPO report World Intellectual Property Indicators 2025 [5], the global number of patent applications reached approximately 3.7 million, representing an increase of 4.9% compared with 2023. Positive dynamics were also observed for utility models (+4.0%) and industrial designs (+2.2%). At the same time, trademark filing activity remained largely stable, showing only a slight decline of around 0.1%.

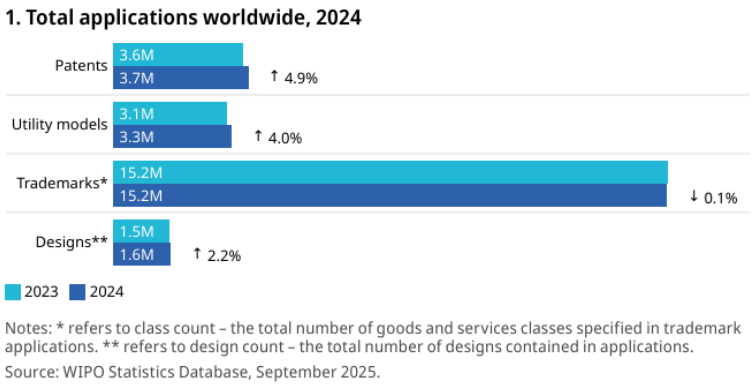


Fig. 2. Total applications worldwide, 2024.

The decrease in trademark activity may be interpreted as reflecting broader macroeconomic uncertainty, which primarily affects sectors associated with consumer markets, branding, and marketing activities. At the same time, the growth in patent and utility model filings indicates the continued strategic focus of companies on the protection of technological solutions and technical innovations.

Thus, despite short-term fluctuations across individual categories of intellectual property rights, the global intellectual property system as a whole continues to demonstrate steady growth in filing activity, reflecting the increasing importance of intangible assets in the modern global economy.

An equally indicative trend is the growing number of active intellectual property rights worldwide. In particular in 2023, the total number of active trademark registrations reached 88.2 million (+6.4%), while the number of registered industrial designs amounted to 6.1 million (+10.5%) and active patents totaled approximately 18.6 million (+7.6%).

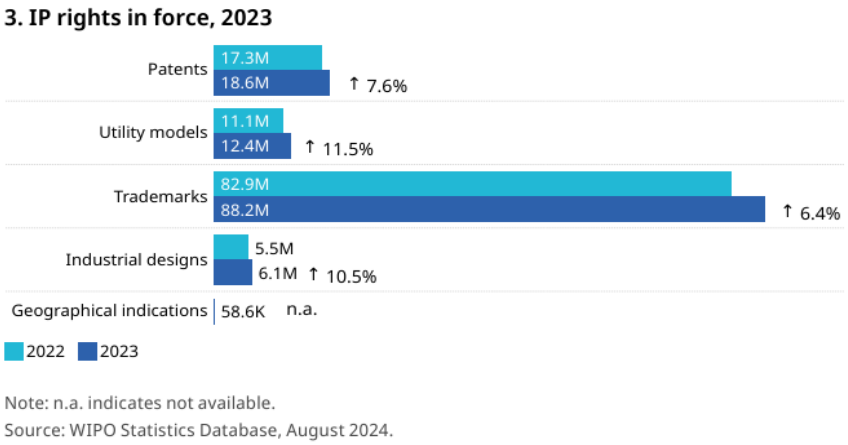


Fig. 3. IP rights in force, 2023

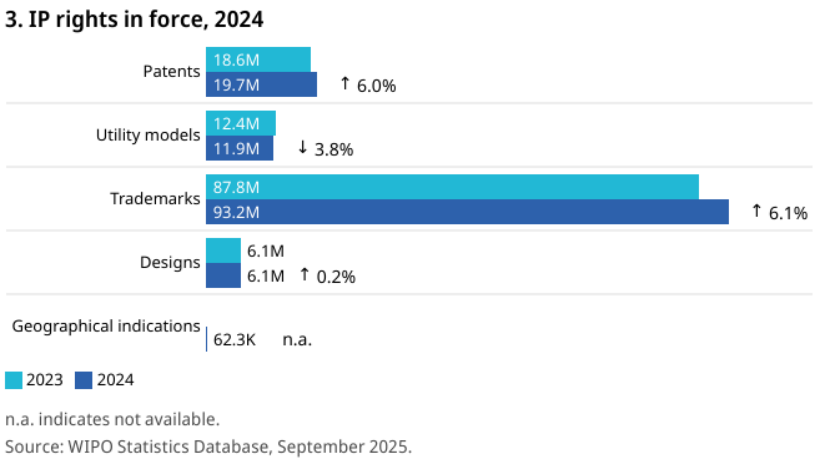


Fig. 4. IP rights in force, 2024

The analysis of statistical data for 2024 [2,4] makes it possible to assess the structure and dynamics of the global stock of intellectual property rights in force. The most widespread

form of intellectual property protection continues to be trademarks, with the total number of active registrations reaching approximately 93.2 million, representing an increase of 6.1% compared with 2023.

The number of patents in force also continued to grow, increasing from 18.6 million in 2023 to approximately 19.7 million in 2024, which corresponds to a rise of about 6.0%. At the same time, the total number of industrial designs in force remained largely stable at around 6.1 million, showing only a marginal increase of about 0.2%.

A somewhat different dynamic can be observed with respect to utility models. Despite their considerable scale, the global stock of active utility model rights declined from 12.4 million to 11.9 million, representing a decrease of approximately 3.8%.

Overall, these figures indicate the continued expansion of the global portfolio of protected intellectual property assets. The dominance of trademarks reflects their role as the most widely used instrument of market differentiation, while the steady growth in patents confirms the sustained importance of technological innovation in the contemporary global economy. These indicators confirm that, despite short-term fluctuations in filing activity, the overall volume of protected intellectual property assets continues to grow.

Thus, a dual trend can be observed. On the one hand, companies appear to be optimizing expenditures related to brand promotion and trademark protection, which is reflected in the relative stabilization of trademark filings. On the other hand, businesses continue to strengthen their positions in technological and production-oriented sectors through patent protection and related instruments of industrial property. This also reflects a gradual shift toward the growing importance of the "invisible" value of assets — technologies, design solutions, and innovative technical developments.

In this context, it becomes relevant to examine how these trends are reflected in the global geographical distribution and regional concentration of intellectual property activity.

3. Regional concentration and distribution of intellectual property activity

The geographical distribution of intellectual property filings demonstrates a pronounced regional concentration of innovation activity. According to WIPO data [1], a significant share of global intellectual property filings is concentrated in Asia. In 2023, applications filed with offices in the Asian region accounted for 68.7% of all patent applications, 69% of trademark applications, 66.7% of industrial design applications, and almost 99% of utility model filings [2].

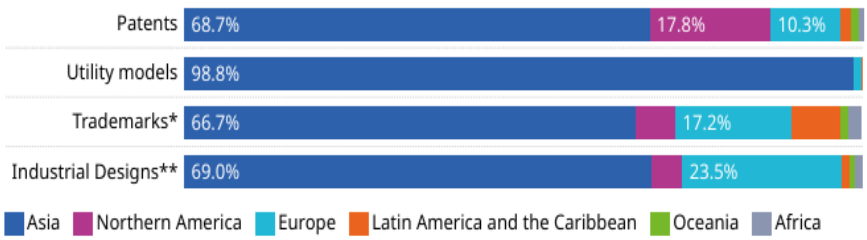


Fig. 5. Information for regional concentration and distribution of intellectual property activity

Data for 2024 confirm the persistence of this pattern. As illustrated, Asia continues to dominate global intellectual property activity[4].

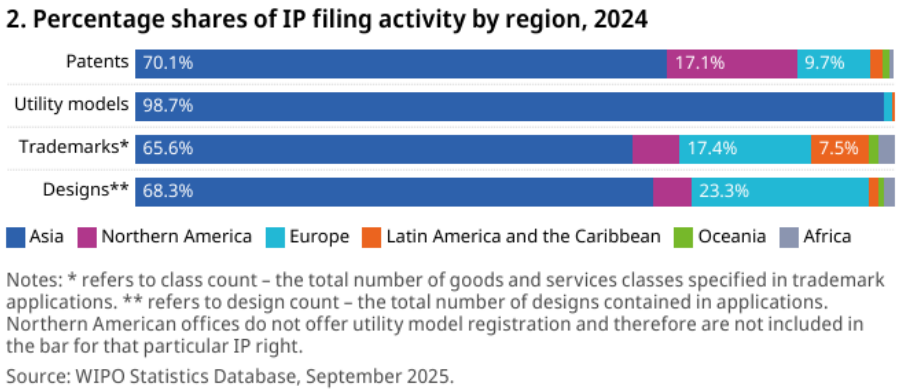


Fig. 6. Percentage shares of IP filing activity by region, 2024.

In 2024, offices in the region accounted for approximately 70.1% of global patent applications, 65.6% of trademark filings, and 68.3% of industrial design applications. The concentration is even more pronounced in the case of utility models, where the share of Asian offices reached 98.7%, reflecting the particularly widespread use of this form of protection in several Asian jurisdictions.

This concentration reflects the rapid development of technological and industrial capacities in a number of Asian economies, as well as the increasing role of the region in the global innovation system. The scale of filing activity indicates that Asia has become the primary center of global intellectual property activity, particularly in sectors related to manufacturing, applied technologies, and product design.

At the same time, North America and Europe continue to maintain strong positions in the field of high-technology innovation and research-intensive industries. In 2024, North America accounted for about 17.1% of patent filings, while Europe represented approximately 9.7% [6]. In the case of trademarks and industrial designs, Europe demonstrates a somewhat larger share, reflecting the strong role of European markets in brand protection and product design.

Other regions - including Latin America and the Caribbean, Oceania, and Africa - account for comparatively smaller proportions of global intellectual property filings. Nevertheless, their participation indicates the gradual expansion of innovation activity beyond the traditional centers of technological development.

Nevertheless, the overall structure of global filing activity remains highly concentrated, which highlights the dominant role of East Asian economies - primarily China, the Republic of Korea, and Japan — in the contemporary architecture of the global intellectual property system.

Data for 2023 indicate a high concentration of filing activity within a limited number of leading intellectual property offices. This pattern largely reflects the dominant role of East Asian economies - primarily China, the Republic of Korea, and Japan — in the contemporary architecture of the global intellectual property system. China continues to play a leading role in these processes. In 2023, the Chinese patent office accounted for approximately 47.2% of all patent applications worldwide, more than 54% of industrial design filings, and nearly half of all trademark applications. Notably, the vast majority of these applications are filed by Chinese residents, which reflects not only the technological maturity of the Chinese economy but also a high level of institutional trust in the national intellectual property protection system. The Republic of Korea and Japan also maintain consistently strong positions in global patent

activity. At the same time, the relative shares of the United States and the European Patent Office in the global distribution of patent filings have gradually declined over the past decade. In particular, the share of the United States decreased from approximately 22.3% of global patent applications ten years ago to about 16.8%, while the share of the European Patent Office fell to around 5.6% [6]. This trend may reflect both the relative maturity and saturation of innovation systems in advanced economies and the growing competitive pressure from countries demonstrating higher filing intensity and rapidly expanding technological capabilities.

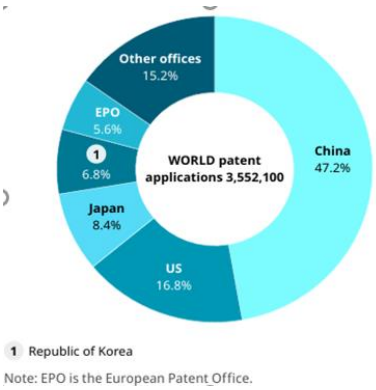


Fig. 7. World patent applications, 2023

Data for 2024 confirm the persistence of this trend. According to WIPO statistics, the five largest offices - China, the United States, Japan, the Republic of Korea, and the European Patent Office - together accounted for approximately 85.5% of all patent applications filed worldwide[3, 5].

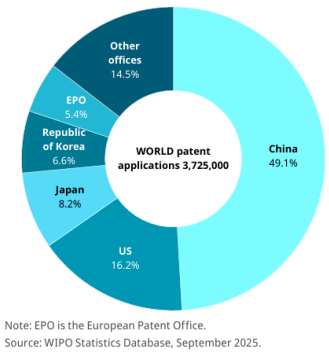


Fig. 8. World patent applications, 2024

China remains the clear leader, accounting for approximately 49.1% of global patent applications, meaning that nearly every second patent application worldwide is filed in China[7]. The United States accounted for about 16.2% of global patent filings, followed by Japan (8.2%) and the Republic of Korea (6.6%), which continue to hold strong positions within the global innovation system. The share of the European Patent Office amounted to approximately 5.4%[6]. Theses further confirm the longer-term trend discussed above, namely the gradual decline in the relative shares of the United States and the European Patent Office in the global distribution of patent filings.

At the same time, the growing concentration of patenting activity among the largest offices has been driven largely by the rapid expansion of patent filings in China. Over the past decade, China's share of global patent applications increased from 34.6% in 2014 to 49.1% in 2024, reflecting the accelerating development of the country's technological and innovation capacity[7]. Thus, the analysis of WIPO statistics indicates a dual concentration of global intellectual property activity. On the one hand, a significant share of global filings is concentrated in Asia, which accounts for around 70% of patent applications worldwide. On the other hand, the majority of patent activity is concentrated in a small group of leading intellectual property offices, which together receive more than 85% of all global filings.

Taken together, these developments indicate a structural shift in the global geography of innovation, with the center of patenting activity increasingly concentrated in East Asia, and particularly in China. Amid these developments, countries with emerging innovation ecosystems are demonstrating increasing activity. In particular, India recorded a 15.7% increase in patent applications in 2023, marking the fifth consecutive year of double-digit growth [2].

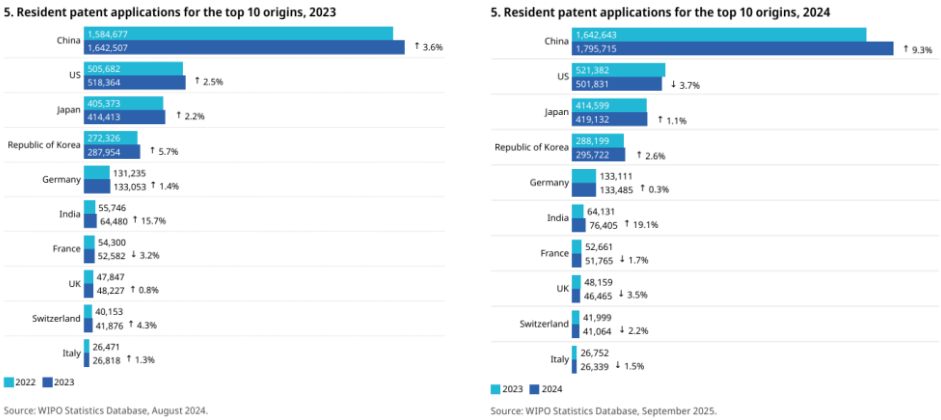


Fig. 9. Resident patens application for the top 10 origins, 2023 and 2024

Data for 2024 further confirm this trend. As illustrated, India increased the number of patent applications to approximately 76.4 thousand, representing a 19.1% increase compared with the previous year. This dynamic indicates that the country is gradually strengthening its position among the major sources of patent applications worldwide and is becoming one of the most rapidly expanding innovation actors.

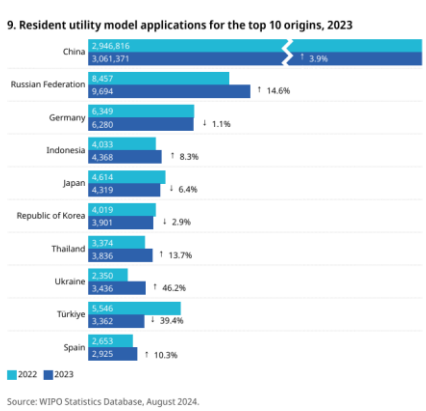
At the same time, the dynamics of patent activity among leading countries remain uneven. China continues to demonstrate strong growth (+9.3%), while the United States recorded a moderate decline (–3.7%). Meanwhile, Japan and the Republic of Korea maintain relatively stable growth (+1.1% and +2.6%, respectively), reflecting the resilience of their national innovation systems. Similar patterns can also be observed in the field of utility model applications, which often serve as a more accessible instrument for protecting technical solutions in developing innovation systems. At the same time, a comparable increase in filing activity in 2023 can be observed in several emerging economies — including Turkey, Viet Nam, Brazil and Ukraine — particularly in the segments of trademarks and industrial designs.

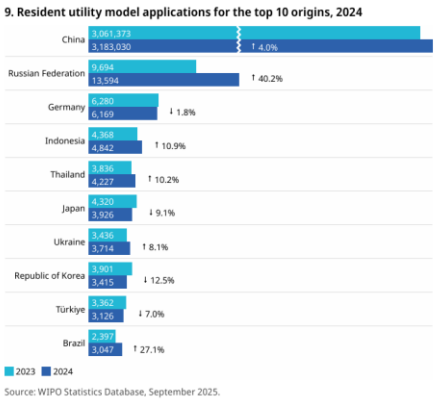
Presented illustrate the distribution of resident utility model applications among the top origins in 2024. As in previous years, China remains the dominant actor in this category, with more than 3.18 million applications, demonstrating a further increase of 4.0% compared with 2023. At the same time, several countries demonstrate particularly notable growth in

utility model filings. The most significant increase is observed in the Russian Federation, where the number of applications rose by 40.2%, reaching approximately 13.6 thousand. A substantial rise is also recorded in Brazil (+27.1%), while Indonesia (+10.9%) and Thailand (+10.2%) show steady growth in the use of this form of protection. Moderate increases are observed in Ukraine (+8.1%), while Türkiye demonstrates a slight decline (–7.0%). At the same time, a decrease in filings is recorded in Japan (–9.1%) and the Republic of Korea (–12.5%), whereas Germany remains relatively stable (–1.8%).

Overall, these s suggest that the use of utility models continues to expand in several emerging innovation systems, where this instrument often serves as a more accessible mechanism for protecting incremental technical solutions.

Thus, the regional shift toward Asia is becoming a stable structural feature of the g lobal intellectual property system. At the same time, the contribution of developing economies to global filing activity continues to grow. This trend creates new challenges for the global harmonization of intellectual property protection standards and highlights the need to take regional specificities into account in the development of international intellectual property policy.



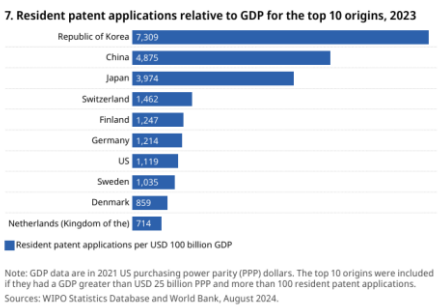


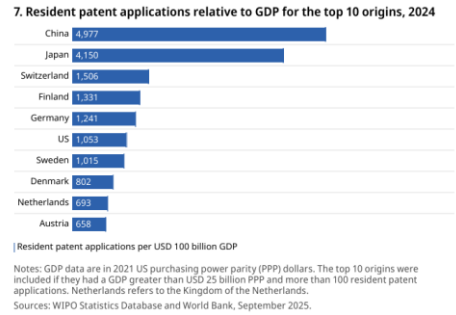
4. Filing intensity relative to GDP

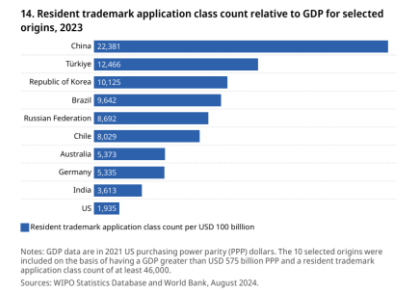
In addition to absolute filing volumes, an important indicator of the maturity of national intellectual property systems is the intensity of filings relative to gross domestic product (GDP)[9]. This metric makes it possible to compare countries with different economic sizes and to identify the relative level of engagement in the creation and protection of intellectual property [10]. According to WIPO statistics, in 2023 the Republic of Korea ranked first in terms of resident patent applications per USD 100 billion of GDP, with 7,309 applications. The second and third positions were occupied by China (4,875) and Japan (3,974). Notably, Finland, surpassing Germany, entered the top five for the first time with 1,247 applications per USD 100 billion of GDP, reflecting a strong concentration of research and development resources [4].

Data for 2024 indicate certain shifts in the distribution of leading countries. China recorded the highest intensity of patent filings with 4,977 applications per USD 100 billion of GDP, followed by Japan (4,150). Among European economies, Switzerland (1,506), Finland (1,331) and Germany (1,241) demonstrate relatively high patent filing intensity, reflecting the strong role of research-intensive and technology-driven sectors in their economies.

A similar analysis for trademarks reveals a somewhat different structure of filing intensity. In 2024, China remains the leading country with 20,198 classes per USD 100 billion of GDP, followed by Türkiye (12,144), Brazil (10,226) and the Russian Federation (8,599). Several smaller economies, including Chile, also show relatively high filing intensity, reflecting the active development of domestic markets and entrepreneurial activity:

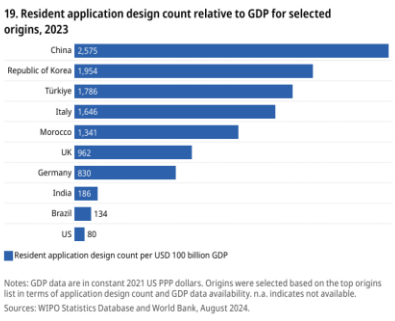


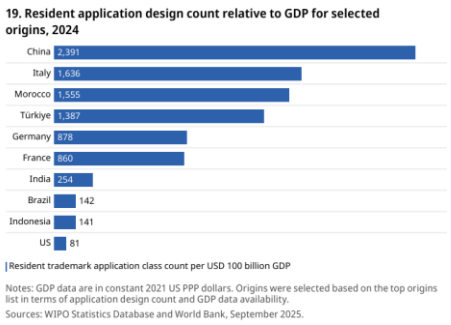






In the field of industrial designs, filing intensity is likewise concentrated in a limited number of economies. According to 2024 data, China (2,391) demonstrates the highest level of design filings relative to GDP, followed by Italy (1,636), Morocco (1,555) and Türkiye (1,387). At the same time, in several large economies such as the United States, this indicator remains relatively low (81 applications per USD 100 billion of GDP), which may reflect differences in industrial structure and national strategies for design protection:





Indicators of filing intensity adjusted for economic size provide valuable insights into the effectiveness of national innovation and intellectual property policies [8]. High values often reflect a combination of strong R&D investment, supportive innovation policies, tax incentives for innovative firms, and well-developed institutional infrastructures for intellectual property protection.

5. Conclusion

The analysis of the latest statistical data published by the World Intellectual Property Organization allows several key conclusions to be drawn regarding the development of the global intellectual property system in 2023–2024. First, the overall level of intellectual property activity remains consistently high across the main categories of industrial property rights. Patent, industrial design and utility model filings continue to demonstrate moderate growth, while trademark activity remains relatively stable despite short-term fluctuations. This confirms the continuing importance of intellectual property as a strategic asset in the modern knowledge-based economy. Second, the geographical distribution of filing activity demonstrates an increasingly pronounced regional concentration in Asia. Countries such as China, the Republic of Korea and Japan maintain leading positions in the global intellectual property system. At the same time, the growing share of China in global patent filings and the strong presence of Asian offices across multiple IP categories indicate a structural shift in the global innovation landscape. Third, the analysis highlights the growing role of emerging and middle-income economies in the global intellectual property system. Countries such as India, Brazil, Indonesia, Thailand and Ukraine demonstrate increasing filing activity, particularly in patents, utility models and industrial designs. This trend reflects the gradual expansion of innovation ecosystems beyond traditional technological centers. Taken together, these developments suggest that the global intellectual property system is undergoing both quantitative expansion and structural transformation. While the overall number of filings continues to grow, the geography of innovation is becoming more diversified, with new regional actors strengthening their positions in the global innovation landscape. These trends create new challenges for the international governance of intellectual property, including the need for greater coordination of national IP policies, the adaptation of regulatory frameworks to rapidly evolving technological sectors, and the integration of intellectual property statistics into broader economic and innovation policy analysis. In this context, intellectual property statistics are becoming an increasingly important analytical tool for understanding global innovation dynamics and shaping evidence-based economic and technological policies.

References

1. World Intellectual Property Organization (WIPO) https://www.wipo.int/en/web/about-wipo
2. IP Facts and Figures 2024 https://www.wipo.int/web-publications/ip-facts-and-s-2024/en/index.html
3. IP Facts and Figures 2025 https://www.wipo.int/web-publications/ip-facts-and-s-2025/assets/85937/943-ip-facts-and-s-2025.en.pdf
4. World Intellectual Property Indicators 2024 https://www.wipo.int/publications/en/details.jsp?id=4759&plang=EN
5. World Intellectual Property Indicators 2025 https://www.wipo.int/edocs/pubdocs/en/wipo-pub-941-17-2025-en-world-intellectual-property-indicators-2025.pdf
6. European Patent Office https://www.epo.org/
7. China National Intellectual Property Administration https://english.cnipa.gov.cn/
8. Unlocking IP-backed Financing - Country Perspectives https://www.wipo.int/publications/en/series/index.jsp?id=241
9. Measuring the Contribution of Intellectual Property to GDP https://etisc.wipo.int/news/measuring-contribution-intellectual-property-gdp
10. Vladlena Lisenco. Intellectual Property in the Innovative Development of the World Economy https://www.intechopen.com/chapters/1122825

A Review on Various Deep Learning Techniques for the Detection of Brain Disorders

Enteamen Yengkhom, Sivakumar Rajgopal

Vellore Institute of Technology, Vellore, Tamil Nadu, India
enteamen.yengkhom2025@vitstudent.ac.in, rsivakumar@vit.ac.in

Abstract. Electroencephalography (EEG) is a method that has established itself for the diagnosis of brain-related and mental disorders along with its ability to provide high temporal resolution**.** This review presents the EEG-based diagnosis of brain disorders, knowledge, and understanding. The research covered the disorders of epilepsy, Alzheimer's disease, ADHD, sleep disorders, schizophrenia, depression, and Parkinson type. These studies were collected from 25 peer-reviewed papers about the use of the following techs: convolutional neural networks (CNNs), long short-term memory networks (LSTMs), transformers, graph convolutional networks (GCNs), spiking neural networks (SNNs), and explainable AI (XAD). Reported accuracies varied from 83% in pediatric sleep staging up to over 98% in seizure detection and Alzheimer's disease classification. Results highlight the growing use of hybrid DL configurations, patient-specific modeling, and interpretable systems - the latter being key to real-time clinical usage of EEG analysis. The range of future studies might encompass privacy-preserving federated learning, multimodal fusion with neuroimaging and biomarkers, and disease progression predictive modeling. This review delineates recent advancement as well as a blueprint for the AI-driven EEG translation to clinical practice.

Keywords. Electroencephalography (EEG), deep learning (DL), seizure detection, neurological disorder diagnosis, explainable artificial intelligence (XAI).

1. Introduction

Electroencephalography (EEG) is among the most popular non-invasive neuroimaging modalities for evaluating brain function and recognizing neural abnormalities. It provides excellent temporal resolution, is relatively cheap and portable, hence being a vital tool for both the clinic and research.

Conventionally, the interpretation of EEG has been dependent solely on the experts' visual checking, which is naturally subjective, time-demanding, and limited to very faint pattern recognizing. The adoption of computational methods that could facilitate in rendering correct, efficient, and objective EEG signal interpretations to the clinicians has been drawn by these limitations.

Indeed, over the last few years, a lot of promise has been shown by the developments in the field of artificial intelligence (AI) and deep learning (DL) for an automated EEG-data analysis. Herein lies a fundamental difference between traditional machine learning and the latter: whereas the former is feature-driven, with most of the features engineered by hand, the latter is feature extraction-driven, where complex hierarchical features can be automatically extracted from raw or minimally processed EEG signals by deep learning architectures such as convolutional neural networks (CNNs), recurrent neural networks (RNNs), long short-term memory networks (LSTMs), and transformers [1–4].

These are just some of the indisputable instances where deep learning methods have been put to work in a very broad spectrum of brain disorders, epileptic syndromes, Alzheimer's disease, attention deficit hyperactivity disorder (ADHD), sleep disorders, schizophrenia, depression, and Parkinson's disease among others [5–10]. The reported performances scope is task and dataset dependent, for example, lightweight pediatric sleep staging systems designed for interpretability and clinical feasibility reach approximately 83% accuracy [16], while seizure detection and Alzheimer's classification models attaining over 99% accuracy [2,6] have been reported.

The recent research has demonstrated the deployment of different and varied methods ranging from hybrid architectures combining CNN and LSTM [6, 19], to sophisticated attention-based frameworks such as transformers [1,3,4,17]. The latter is noted for high interpretability, which is very valuable in attaining clinical trust. The utilization of the explainable artificial intelligence (XAI) methods, such as LIME and SHAP, is just another example which reflects the place of interpretability not only in scientific discourse but more importantly in the clinical trust [5, 25]. Multimodal fusion, one where EEG can be combined with other bio signals or sensors to achieve a more robust diagnosis, is another potential area that is gaining traction [23].

Despite these advancements, there still exist some problems such as small sizes of datasets, class imbalance, poor generalization across patient populations, and the lack of longitudinal studies that capture disease progression.

The current review is an attempt to amalgamate the data related to the diagnosis of brain disorders based on EEG published between 2020 and 2025 while using deep learning methods. The review investigates 25 peer-reviewed research papers ranging over epilepsy, Alzheimer's disease, ADHD and other cognitive disorders, sleep disorders, schizophrenia, depression and Parkinson's disease. It not only helps in the comprehension of EEG-based deep learning for brain disorder diagnosis by the academic community and clinicians but at the same time also points out the areas of research that require more studies and the future directions those studies may take by summarizing recent trends, showcasing methodological breakthroughs, and pinpointing research gaps in the field.

2. Definition of Concepts

2.1 Electroencephalography (EEG)

Electroencephalography (EEG) is a non, invasive method that records the brain's electrical signals via electrodes that are attached to the scalp. The main characteristics of EEG are high temporal resolution, low cost, and the potential of capturing both transient and long, term brain dynamics. The application of EEG signals is very broad in the medical diagnostics field of the diseases of the nervous system and the mind, which are the case of epilepsy, Alzheimer and sleep disorders among others. However, raw EEGs are typically noisy, high dimensional and non-stationary, which leads to the difficulty of experts in decoding them and thus computationally assisted methods have been implemented as a solution to this problem [1, 7].

2.2. Artificial Intelligence and Deep Learning

Artificial intelligence (AI) is a term used for the capability of a computer or a computer, controlled robot to do tasks that are usually done by humans, which include decision making, pattern recognition, and so on. Deep learning (DL) models have gained a lot of traction in the field of EEG analysis and have become the leading ones within the total AI framework. One of these models is the convolutional neural networks (CNNs) which is efficient in drawing the

spatial patterns while the other is the recurrent neural networks (RNNs) and long short, term memory networks (LSTMs) which are designed to extract the temporal dynamics [6, 19]. Lately, attention, based paradigms like transformers have been integrated for far, reaching sequence modeling in EEG [1, 4, 17] and brain connectivity studies have propelled the use of graph convolutional networks (GCNs) [8, 13]. The diversity of methods has been further broadened by the inclusion of spiking neural networks (SNNs) and explainable AI (XAI) frameworks [14, 25].

2.3 Clinical Disorders in Focus

The first question asks for the clinical conditions that appear in this review and represent areas of research based on the diagnostic use of EEG.

- Epilepsy: It is a disease where abnormal neuron discharges are the cause of repeated seizures besides that EEG is used as a gold standard in seizure detection and prediction [2, 3, 5].
- Alzheimer's Disease: A neurodegenerative disorder characterized with gradual decline in cognitive abilities. EEG changes like slowing rhythms and interrupted connectivity can act as biomarkers [6, 9].
- Attention Deficit Hyperactivity Disorder (ADHD): It is a neurodevelopmental disorder whose main symptoms are inattention and hyperactivity. Some EEG abnormalities that have been found include changed spectral power and disrupted connectivity [11, 15].
- Sleep Disorders: The above, mentioned diseases are studied by polysomnography (PSG), which main stages of the sleep are carried out by EEG [16, 18].
- Schizophrenia and Depression: Mental diseases that differ from each other only by the presence of some subtle but detectable EEG biomarkers that are the changes of oscillatory patterns and connectivity [19, 21].
- Parkinson's Disease: It is a disorder of the central nervous system that affects movement, the symptoms can be very helpful EEG features will be clearer when you combine the detected features with the help of wearable sensor data [22, 23].

Once these key concepts have been identified, the stages for an organized review of the present research trends, methodologies and results across these disorders are established.

3. Current Research Trends

3.1 Epilepsy and Seizure Detection

Just with the advent of attention, based models, epileptic seizure prediction has made substantial headway. Reporting accuracies over 95% for transformer designs applied to benchmark scalp EEG datasets, Yan et al. [1] and Bhattacharyya et al. [2], both achieved striking performance comparable to or better than conventional recurrent models. Zhu et al. [3] elevated that level further by blending multidimensional transformers with the recurrent neural network, which allowed him to exceed 96% of accuracy. By coupling raw time series data together with spectrogram features using a dual, modality transformer, Qin et al. [4] not only increased the accuracy to more than 97% but also gained more robustness for cross, patient experiments. Reaching roughly 94% of accuracy while also promoting interpretability, Ahmad et al. [5] fused efficient feature selection with an explainable classifier. Basically, these results collectively depict the striving for highly accurate (>94%) seizure detection pip elines

of which transformers and hybrid methods dominate the field by setting the bar for new benchmarks.

3.2 Alzheimer's Disease

Among the various disorders in the review, EEG, based Alzheimer's classification retains some of the highest accuracies. DeepADNet (CNN, LSTM) proposed by Ho et al. [6] attained over 98% accuracy across different stages of the disease. AlSharabi et al. [7] leveraged empirical mode decomposition combined with deep learning and achieved approximately 95% accuracy in clinical cohorts. Shan et al. [8] utilized a spatial, temporal GCN on EEG, derived connectivity, and the accuracy ranged from 92% to 96%, while Cao et al. [9] supported with a deep pyramid CNN showed a performance of over 97%. Kachare et al. [10] presented a lightweight CNN (LCADNet) that could efficiently maintain the balance with the accuracy close to 94%. These results indicate that EEG classification of Alzheimer's is highly feasible if most of the approaches are capable of more than 90% accuracy.

3.3 ADHD and Cognitive Disorders

The detection of ADHD has been gradually progressing, but it still generally lags the results of the seizure and Alzheimer's research fields. Altun et al. [11] created CNNs with scalograms and reported the accuracy in the range of 85, 88%. Nonlinear features and wavelet transforms were used by Joy et al. [12], and they were able to achieve about 87% of the accuracy. Li et al. [13] made a step forward toward graph convolutional networks and claimed 90% of accuracy with multi, domain features. Baibulova et al. [14] stressed that one could keep the classification accuracy at the level near 89% while focusing on interpretability. Alhussen et al. [15] utilized NeuroDCT, ICA with AttentionNet optimization and attained the accuracy of more than 91% at the early stage of ADHD detection. The overall scope of accuracies for studies related to ADHD is from 80% to 92%. Deep learning, although offering substantial improvements, shows that more significant gains might be possible only with larger datasets and multimodal inputs.

3.4 Sleep Disorder

EEG remains the principal method of research in sleep staging and the operational range largely depends on the variety of models. Zhu et al. [16]79] built a lightweight edge AI model for sleep staging in children using single, channel EEG and the accuracy obtained was 83%. This was in support of their goal to enable deployment in low, resource environments. Fox et al. [17] demonstrated that a full, night polysomnography, trained transformer got better results (~91% accuracy) as it could capture the long, range sleep dynamics. The best results of the market were obtained by Lee et al. [18] with an explainable vision transformer on multimodal PSG at the level of over 92%. Besides providing clinically relevant visual explanations, this development clearly shows the trade, off that exists between lightweight edge systems and high, performing but computationally heavier transformer, based architectures.

3.5 Schizophrenia and Depression

The EEG biomarkers of psychiatric disorders are less obvious thus; the accuracies are modest but still encouraging. The CNN, LSTM models reported by Shoeibi et al. [19] were able to achieve an accuracy of around 87% for schizophrenia diagnosis. The attention, based EDT model created by Ying et al. [20] for depression got the accuracy in the range of 88, 90%.

The application of multichannel CNN to schizophrenia EEG resulted in ~85% of accuracy for the differentiation of patients from the control group [21]. Although these numbers are lower than those for seizure or Alzheimer's detection, they nonetheless confirm the promise of deep learning in psychiatric domains where diagnostic hints are few and faint.

3.6 Parkinson's Disease and Multi, Disorder Diagnosis

With multimodal data, the Parkinson's disease studies are a different story. Li et al. [22] applied a CNN, GNN hybrid to EEG, and the outcome was the approximate 91% of accuracy. Zhang et al. [23] enrolled EEG together with data from wearable sensors and thus, substantially, the performance made a big jump and was over 93%. In brief, the findings suggest that brain activity alone may not fully depict Parkinson's illness however, combining different modalities can aid in the accuracy of the diagnosis.

Table 1. Comparative Summary of Reviewed Studies

Disorder	Dataset (if reported)	Methodology / Model	Reported Accuracy / Findings	Ref.
Epilepsy/ Seizures	CHB-MIT, scalp EEG (various)	Transformer models [1,2]; Transformer + RNN fusion [3]; Dual-modality transformer [4]; Feature selection + explainable classifier [5]	>95% (up to ~97% in hybrid/dual-modality models)	[1–5]
Disorder	**Dataset (if reported)**	**Methodology / Model**	**Reported Accuracy / Findings**	**Ref.**
Alzheimer's Disease	Resting-state EEG (clinical)	CNN-LSTM hybrid (DeepADNet) [6]; EMD + DL [7]; Spatial–temporal GCN [8]; Deep Pyramid CNN [9]; Lightweight CNN (LCADNet) [10]	90–99% (DeepADNet and pyramid CNN >97%)	[6–10]
ADHD & Cognitive Disorders	Clinical ADHD EEG datasets	Scalogram + CNN [11]; Non-linear + wavelet [12]; GCN multi-domain [13]; Interpretable hybrid DL [14]; AttentionNet with NeuroDCT-ICA [15]	80–92% (best ~91% with AttentionNet optimization)	[11–15]

Sleep Disorders	Paediatric and adult PSG datasets	Lightweight edge AI [16]; Full-night transformer [17]; Explainable vision transformer on multimodal PSG [18]	83% (lightweight paediatric) to>92% transformer-based)	[16–18]
Schizophrenia & Depression	Multichannel EEG (clinical)	CNN-LSTM [19]; Attention-based EDT [20]; CNN for schizophrenia [21]	~85–90% across disorders	[19–21]
Parkinson's Disease	EEG + wearable sensor fusion	CNN-GNN hybrid [22]; EEG + wearable fusion model [23]	~91% (EEG-only) to >93% (EEG + wearable fusion)	[22–23]
Methodology / General	—	Systematic review [24]; XAI with LIME/SHAP for seizure detection [25]	~94% accuracy with explainability integration. Highlighted data scarcity, interpretability, need for trust	[24–25]

4. Models and Methodologies in EEG-Based Diagnosis

4.1 Convolutional Architectures

Convolutional neural networks (CNNs) are still the major components of EEG, based classification, especially for Alzheimer's disease and movement disorders. The development of CNNs has brought about commendable results, with Alzheimer's disease diagnosis often scoring between 90% and 99% accuracy [6, 9, 10]. Pyramid CNNs [9] got around 97% while light CNNs like LCADNet [10] kept the good performance with about 94% accuracy. Besides, CNNs applied to schizophrenia [21] and Parkinson's disease [22] have also made a breakthrough where the performance accuracies ranged 85, 91%. These performances point to CNNs as dependable spatial feature extractors yet, their capacity may reach a ceiling without mixing with other models.

4.2 Recurrent and Sequential Models

Recurrent models long short-term memory networks (LSTMs), are good at representing the time, varying dependencies of EEG signals. CNN, LSTM combinations like DeepADNet used for Alzheimer's [6] and schizophrenia [19] classification have been found to be consistently accurate in the range of ~87, 98%, thus they surpassed single, model baselines. Their effectiveness is due to the coupling of CNN spatial features with LSTM's capability of following sequential EEG variations. On the other hand, RNN, based architectures may be

slow when it comes to running heavy computations and they might not be as scalable as transformer architectures.

4.3 Attention Based Transformers

The reason why transformers have quickly evolved seizure detection, ADHD diagnosis, and sleep staging is that they can handle long dependencies efficiently. High accuracies above 95% have been the norm in seizure prediction literature [1, 4], with one of the best performances of 97% achieved by a dual, modality transformer [4]. Attention, based models for ADHD scored >91% [15] and transformer, based sleep staging stayed over 90% [17, 18]. These results indicate that transformers are setting new levels of performance across different disorders, although their high computational cost still makes it tricky for real, time clinical use.

4.4 Graph Based Models

Graph neural networks (GNNs) and graph convolutional networks (GCNs) use the brain's connectivity patterns to enhance classification. Shan et al. [8] achieved 92, 96% in Alzheimer's detection by implementing a spatial, temporal GCN, and Li et al. [13] reported 90% in ADHD detection by utilizing multi, domain graph features. Also, CNN, GNN hybrids have helped Parkinson's diagnosis with the accuracy reaching up to 91% [22]. These results show network level biomarkers can be highly relevant to clinics, particularly in diseases related to disrupted connectivity.

4.5 Time, Frequency and Feature Engineering Approaches

Feature extraction is still needed when deep learning alone is not enough. The use of scalograms with CNNs for ADHD detection led to an accuracy of about 85, 88% [11] while nonlinear + wavelet features reached 87% [12]. Empirical mode decomposition (EMD) combined with deep models was given an accuracy of about 95% for Alzheimer's [7]. The development of feature, engineered approaches might lag the state-of-the-art deep models, but they are still significant for the sake of dataset size and model interpretability.

4.6 Explainable AI (XAI)

Interpretability is a key aspect of clinical trust. For example, Ahmad et al. [5] obtained nearly 94% accuracy in seizure detection while maintaining feature, level explanations. In sleep staging, vision transformers [18] gave over 92% accuracy together with saliency maps for assisting clinical validation. In their work, Khan et al. [25] added LIME and SHAP to seizure detection pipelines, thus the system was able to get about 94% of accuracy with interpretable predictions. These studies collectively provide evidence that explainability could be implemented with little accuracy tradeoff, thereby increasing confidence in AI based diagnosis.

4.7 Multimodal and Edge Friendly Systems

Efficiency and practicality are other advantages that are coming more into the limelight. For instance, Zhu et al. [16] created a child sleep staging system with roughly 83% accuracy made possible using single, channel EEG and edge AI, proving its great potential in places with limited resources. Moreover, the combination of EEG and wearable sensors led to a rise in the detection of Parkinson's disease to over 93% [23] which was higher than that of EEG only studies. The importance of such strategies lies in the equilibrium between model complexity, computational efficiency, and clinical deployability.

5. Discussion and Challenges

The reviewed studies collectively demonstrate the rapid evolution of EEG, based deep learning for brain disorder diagnosis. Various key points come up when studying the methodologies, the results, and the clinical use cases.

5.1 Performance Across Disorders

The most researched diseases based on the EEG method are still epilepsy and Alzheimer's disease, for which several transformers, based and hybrid CNN, LSTM models have reported the best part of their accuracy above 95% [2, 4, 6, 9]. On the other hand, the study of ADHD, schizophrenia, and depression has made less progress and the accuracy is usually between 80 and 90 % [12, 14, 20]. The difference between these two groups is connected both to the subtlety of EEG biomarkers in psychiatric conditions and to the smaller sizes of the available datasets. The focus of sleep staging models is, above all, on computational efficiency, especially in the case of pediatric populations, so sometimes the accuracy is sacrificed for deployability [16]. The studies on Parkinson's disease have become increasingly inclined towards multimodal fusion to make up for the limited diagnostic power of EEG alone [22, 23].

5.2 Methodological Trends

Many methodological changes come to light in the review summarized in the paper. Firstly, attention, based models and especially transformers are gaining ground very rapidly and this is the reason why we find them in the forefront of the seizure prediction and the sleep analysis fields. The main property that is highlighted is their capability to handle very long sequences of data, which is a very common case in EEG data processing [1, 3, 17]. Secondly, the use of the graph, based method is becoming more prominent, for example in brain connectivity and alterations of connectivity in Alzheimer's disease and ADHD [8, 13]. Third, putting the interpretability into the spotlight is an imperative issue, not an option anymore; hence, the use of AI models with explainability features that are not a retrofit but are really integrated in the design of the model, like in the case of attention, based annotation with SHAP/LIME [5, 25] and vision transformers for sleep staging [18].

5.3 Data and Generalization Issues

Most of the studies still rely on small or, at best, medium, size datasets collected from a single site, although a lot of methodological advances have been achieved. This is the main reason that their generalizability is very limited. Without tuning the model performance to each specific patient, the accuracy reported is generally not as high. However, doing that reduces the applicability of these models to new populations. Not many studies have done a cross, dataset validation, which is a crucial step for the clinical use of models. Furthermore, the problem of class imbalance still exists and is more noticeable in, for instance, multi, class Alzheimer's disease classification [6, 10] as well as seizure prediction tasks, wherein the number of preictal events is small, thus leading to data sparsity [1, 4].

5.4 Computational and Practical Constraints

The various transformer could indeed be at the forefront of the highest, performing models yielding excellent accuracy; however, their heavy computational complexity is one of the stumbling blocks for the deployment of the models in clinical environments where resources are limited. Users of LCADNet [10] and edge, supported sleep staging models [16]

that are lightweight in terms of architecture may somehow bridge this space, yet several compromises should be accepted in terms of predictive capabilities. Efforts to strike a balance between performance and efficiency are still ongoing and this is particularly the case for real, time applications which represent a serious challenge.

5.5 Clinical Translation and Trust

For the clinical implementation to become a reality apart from accuracy, the factors of interpretability, robustness, and smooth integration into the existing workflows, etc. are very important. Even though the use of explainable AI methods [5, 18, 25] makes a step toward visibility, still, these outputs should be reconfirmed for being accurate reflection. Plus, the ethical concerns like the privacy of patient data in Internet of Medical Things (IoMT) scenarios, to name one, are still undeveloped. Few studies have taken account of evolutions in federated learning or privacy, preserving AI research.

6. Conclusion

This review involved 25 peers, reviewed studies published between 2020 and 2025 that had deep learning as the focus for EEG, based brain disorder diagnosis. The summary of the data shows that epilepsy along with Alzheimer's disease are the most studied disorders to date, while several transformer, based as well as hybrid CNN, LSTM models have also reported their accuracies to be above 95%. On the other hand, research related to psychiatric conditions, for instance, ADHD, schizophrenia, and depression, is still at a very initial stage with low diagnostic performance that can be explained by both the complexity of their biomarkers and the limited availability of large datasets. Sleep disorder studies focus on lightweight and edge, friendly models. At the same time, Parkinson's disease research is progressively depending on multimodal fusion with wearable sensors.

From a methodological perspective, the field is undergoing a transition embracing three major directions: (i) attention, based architectures that capture long, range temporal dependencies, (ii) graph, based models that exploit connectivity information, and (iii) explainable AI frameworks that provide transparency for clinical users. However, despite these improvements, challenges and difficulties remain in terms of the diversity of datasets, class imbalance, generalizability of models, and the need for computationally efficient solutions. The process of moving into clinical practice also entails the necessity of more robustness, privacy, preserving collaboration, and validation of the interpretability methods.

In the future, researchers should focus on using federated learning as the priority which is the technique that facilitates multi, site data sharing without compromising patient privacy, followed by multimodal integration of EEG with imaging and physiological sensors, and finally longitudinal modeling of disease progression. These steps would not only speed up the deployment of AI, driven EEG analysis in real, world clinical settings but also, eventually, they would elevate diagnostic accuracy, foster clinical decision, making and contribute to patient health and well-being.

References

1. J. Yan, J. Li, H. Xu, Y. Yu, and T. Xu, “Seizure prediction based on Transformer using scalp electroencephalogram,” Applied Sciences, vol. 12, no. 9, p. 4158, 2022, doi: 10.3390/app12094158.
2. A. Bhattacharyya, R. B. Pachori, et al., “Epileptic seizure prediction using deep transformer model,” International Journal of Neural Systems, vol. 32, no. 2, p. 2250005, 2022, doi: 10.1142/S0129065722500052.

3. R. Zhu, W. Pan, J. Liu, J. Shang, et al., "Epileptic seizure prediction via multidimensional transformer and recurrent neural network fusion," Journal of Translational Medicine, vol. 22, p. 895, 2024, doi: 10.1186/s12967-024-05678-7.
4. J. Qin, et al., "Dual-modality Transformer with time-series imaging for robust epileptic seizure prediction," Applied Sciences, vol. 15, no. 3, p. 1538, 2025, doi: 10.3390/app15031538.
5. I. Ahmad, et al., "An efficient feature selection and explainable classification method for EEG-based epileptic seizure detection," Computers in Biology and Medicine, vol. 166, p. 107678, 2024, doi: 10.1016/j.compbiomed.2023.107678.
6. T. K. K. Ho, Y. Jeon, E. Na, et al., "DeepADNet: A CNN-LSTM model for the multi-class classification of Alzheimer's disease using multichannel EEG," Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring, vol. 14, no. 1, p. e12345, 2022, doi: 10.1002/dad2.12345.
7. K. AlSharabi, Y. B. Salamah, M. Aljalal, A. M. Abdurraqeeb, and F. A. Alturki, "EEG-based clinical decision support system for Alzheimer's disorders diagnosis using empirical mode decomposition and deep learning techniques," Frontiers in Human Neuroscience, vol. 17, p. 1190203, 2023, doi: 10.3389/fnhum.2023.1190203.
8. X. Shan, J. Cao, S. Huo, L. Chen, P. G. Sarrigiannis, and Y. Zhao, "Spatial–temporal graph convolutional network for Alzheimer classification based on brain functional connectivity of EEG," Human Brain Mapping, vol. 43, no. 17, pp. 5194–5209, 2022, doi: 10.1002/hbm.25994.
9. J. Cao, et al., "Identification of Alzheimer's disease brain networks based on resting-state EEG signals using deep pyramid CNN," Heliyon, vol. 11, no. 3, p. e14858, 2025, doi: 10.1016/j.heliyon.2023.e14858.
10. P. Kachare, D. Puri, S. B. Sangle, I. Al-Shourbaji, A. Jabbari, et al., "LCADNet: A novel light CNN architecture for EEG-based Alzheimer disease detection," Physical and Engineering Sciences in Medicine, vol. 47, no. 2, pp. 445–457, 2024, doi: 10.1007/s13246-024-01425-w.
11. S. Altun, et al., "Automatic diagnosis of attention-deficit hyperactivity disorder using EEG by means of scalogram and deep learning," Clinical Psychopharmacology and Neuroscience, vol. 20, no. 4, pp. 715–728, 2022, doi: 10.9758/cpn.2022.20.4.715.
12. R. C. Joy, et al., "Detection and classification of ADHD from EEG signals using non-linear features and tunable Q-wavelet transform," Computational Intelligence and Neuroscience, vol. 2022, p. 3590973, 2022, doi: 10.1155/2022/3590973.
13. L. Li, X. Guo, Z. Yang, Y. Zhao, X. Liu, et al., "ADHD detection from EEG signals using graph convolutional network based on multi-domain features," Frontiers in Neuroscience, vol. 19, p. 1561994, 2025, doi: 10.3389/fnins.2025.1561994.
14. M. Baibulova, et al., "Interpretable hybrid deep learning framework for classifying ADHD using EEG," Algorithms, vol. 18, no. 8, p. 453, 2025, doi: 10.3390/a18080453.
15. A. Alhussen, A. I. Alutaibi, S. K. Sharma, A. R. Khan, F. Ahmad, G. G. Tejani, et al., "Early ADHD detection with NeuroDCT-ICA and RhinoFish optimization algorithm based AttentionNet," Scientific Reports, vol. 15, p. 6967, 2025, doi: 10.1038/s41598-025-90649-1.
16. L. Zhu, et al., "A lightweight automatic sleep staging method for children using single-channel EEG based on edge artificial intelligence," World Wide Web, vol. 24, no. 4, pp. 1143–1160, 2021, doi: 10.1007/s11280-021-00983-3.
17. B. Fox, et al., "A foundational transformer leveraging full-night PSG encodings for improved sleep stage classification," Scientific Reports, vol. 14, p. 11223, 2024, doi: 10.1038/s41598-024-67890-4.
18. H. Lee, Y. R. Choi, H. K. Lee, J. Jeong, J. Hong, H. W. Shin, and H. S. Kim, "Explainable vision transformer for automatic visual sleep staging on multimodal PSG signals," npj Digital Medicine, vol. 8, no. 1, p. 55, 2025, doi: 10.1038/s41746-024-01378-0.
19. A. Shoeibi, D. Sadeghi, P. Moridian, N. Ghassemi, J. Heras, R. Alizadehsani, et al., "Automatic diagnosis of schizophrenia in EEG signals using CNN-LSTM models," Frontiers in Neuroinformatics, vol. 15, p. 777977, 2021, doi: 10.3389/fninf.2021.777977.
20. M. Ying, X. Shao, J. Zhu, Q. Zhao, X. Li, and B. Hu, "EDT: An EEG-based attention model for feature learning and depression recognition," Biomedical Signal Processing and Control, vol. 93, p. 106182, 2024, doi: 10.1016/j.bspc.2024.106182.
21. "Identification and diagnosis of schizophrenia based on multichannel EEG and CNN deep learning model," Schizophrenia Research, vol. 271, pp. 28–35, 2024, doi: 10.1016/j.schres.2024.07.015.

22. C. Li, X. Huang, X. Chen, et al., "EEG-based Parkinson's detection using CNN-GNN hybrid models," IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 31, no. 5, pp. 1023–1034, 2023, doi: 10.1109/TNSRE.2023.3245678.
23. Y. Zhang, et al., "Fusion of EEG and wearable sensor data for Parkinson's disease diagnosis using deep learning," Frontiers in Aging Neuroscience, vol. 16, p. 112233, 2024, doi: 10.3389/fnagi.2024.112233.
24. Y. Roy, H. Banville, I. Albuquerque, A. Gramfort, T. H. Falk, and J. Faubert, "Deep learning-based electroencephalography analysis: A systematic review," Journal of Neural Engineering, vol. 16, no. 5, p. 051001, 2019, doi: 10.1088/1741-2552/ab260c.
25. F. A. Khan, et al., "Explainable AI for epileptic seizure detection in IoMT using LIME and SHAP," Computers in Biology and Medicine, vol. 165, p. 107471, 2024, doi: 10.1016/j.compbiomed.2023.107471.

Machine Learning and Deep Learning for Cuff-less Blood Pressure Monitoring: A Review of Advances and Future Directions

Sneha Majumdar, Sivakumar Rajgopal

Vellore Institute of Technology, Vellore, Tamil Nadu, India
sneha.majumdar2025@vitstudent.ac.in, rsivakumar@vit.ac.in

Abstract. Blood pressure is known to be the leading cause of cardiovascular disease. However, devices based on cuffs only provide measurements from time to time and what is more, these readings are often inaccurate. A new development in IT technology, that is ML and DL, has already allowed the creating of the continuous, cuff-less, and non-invasive monitoring systems. Several ML methods, including Random Forest, Gradient Boosting, and SVR have been successfully used for physiological signals (ECG, PPG, and ABP) to meet global standards. One large-scale Random Forest study (~30 million beats) is a perfect example in which the performance was very strong when a subject-personalized model was used. Some deep learning architectures, e.g. CNNs, LSTMs, BiLSTMs, and hybrid CNN–LSTMs, are capable of automatic feature extraction from various types of signals, and even from video-based rPPG, through certain means like clustering-based preprocessing, PBS fusion, and Conv-LSTM have succeeded in reaching the accuracy level required for clinical application, and in fact, they are also rated as Grade A by BHS. Such advancements are very attractive in the context of wearable, real-time monitoring that can detect the occurrence of nocturnal hypertension, among other hidden risks, an example being a device based on these novel technologies. But along with benefits come certain issues, such as motion artifacts, population-wide generalizability, and calibration-free accuracy, which are picked up as sources of problems needing multimodal sensor fusion and personalized hybrid models to become clinically adopted solutions.

Keywords. Blood, Pressure, Monitoring, Cuff-less, Measurement, Machine Learning (ML), Deep Learning (DL), Physiological Signals (ECG), PPG, ABP, rPPG), Wearable Devices, Multimodal Sensor Fusion

1. Introduction

Blood pressure (BP) is probably one of the most instrumental physiological measurements that depicts the status of the cardiovascular system. It is essential to regularly check BP to be able to detect and control such disorders as hypertension, diabetes, and heart failures. Usually, blood pressure measurement is performed by such methods as auscultatory and oscillometric which rely on an inflatable cuff.

In general, these methods are accurate, but they are also inconvenient, only allow intermittent measurements, and cannot be used for real-time monitoring. On the other hand, continuous blood pressure monitoring provides a way for the earliest detection of hemodynamic changes, thus improving patient outcomes and allowing preventive care.

Wearable sensors and computational intelligence that are the areas of recent advances have been instrumental in changing the paradigm of cuff-less BP monitoring that is now possible with biosignals such as photoplethysmography (PPG), electrocardiography (ECG),

and arterial blood pressure (ABP) waveforms. In particular, PPG has turned into a strong signal because of its ease of use, ability to be carried anywhere, and being inexpensive. Nevertheless, the nonlinear relationship between PPG and BP is difficult and affected by several base and surrounding factors. The problem has led to the implementation of Machine Learning (ML) and Deep Learning (DL) methods to create a model for the complex mapping that links the PPG signal features with the BP values.

Need for Intelligent BP Estimation

ML and DL models can identify that complexity in the interrelations and nonlinearities in biosignals, which, by traditional regression-based models, cannot be accounted for. In addition, they can learn from extensive datasets and apply what they have learned to new situations, which makes them perfect for personalized and continuous BP monitoring.

The deployment of these models together with wearable gadgets, edge computing, and remote monitoring systems is a great innovation in the field of digital health and telemedicine bringing the promise of cuff-less BP estimation being highly accurate and feasible in real- life situations.

2. Definition of Core Concepts

2.1 Photoplethysmography (PPG)

In principle, photoplethysmography evaluates transmitted light through the small blood vessels near the skin's surface, so it is considered a cuffless approach for blood pressure estimation.A little LED sends light into your skin.The amount of light reflected or transmitted through the tissue is detected by a photodetector which notes changes in brightness.Essentially, that rhythmic volume variation of each heartbeat picked up through the photoplethysmogram tells you how blood flows through your circulatory system.

Features Derived from PPG:

Measurements taken over time - like how long a pulse takes to travel, its duration, alongside heartbeats per minute - fall into the category of time-domain features.Looking at how signals change - their energy across different frequencies, likewise their patterns using wavelets - gives us key characteristics.

A pulse's shape reveals itself through its high points - when the heart squeezes, then relaxes - a small dip between them, likewise the total space beneath the waveform.You can gauge characteristics using three measurements: how rigid something is (Stiffness Index), how well it bounces back (Reflection Index), alongside a comparison of pulse strength (Pulse Intensity Ratio).

Advantages: Tiny, they slip right in without fuss - also, they don't break the bank.Ideal for devices you wear or that track from a distance.These tools work well alongside machine learning - even advanced deep learning - allowing for ongoing, automatic blood pressure tracking.**Challenges:** Movement blurs the images, likewise does ambient brightness.How a signal behaves can shift due to age, warmth, likewise complexion.

2.2 Artificial Intelligence (AI), Machine Learning (ML), and Deep Learning (DL)

Computers showing human-like smarts - reasoning, seeing things, choosing what to do - fall under Artificial Intelligence. Within this field, machine learning likewise deep learning play key roles in figuring out blood pressure.Machine Learning (ML)Algorithms learn from data - they spot what usually happens through careful calculationsModels that predict based on relationships include straightforward linear methods, also more complex random forest approaches, likewise gradient boosting techniques.Grouping tools - specifically K-Means - assist when data spreads out or characteristics shift.Sometimes, Support Vector Machines

really shine when data isn't a straight line – instead of simple connections, they handle curves beautifully.

Deep Learning (DL) utilizes multiple layers so it can independently discover what's important within data.CNNs work by pinpointing where things happen within a pulse signal. They're built to grab those specific details.Heartbeats unfold over time, a pattern RNNs alongside LSTMs help us decipher. They reveal how each beat connects to what came before.

A BiLSTM learns patterns by looking at sequences forwards then backwards - essentially considering context from every angle. Consequently, it grasps time-based details more thoroughly.These designs blend convolutional neural networks - good at spotting patterns – with long short-term memory networks that excel at understanding sequences. Essentially, one part finds what's important, while the other figures out how things change over time.Focusing on key parts of pulse signals - rather than everything at once - helps forecasts become more reliable.

2.3 Clinical Disorders in Focus

1). Hypertension (High Blood Pressure):-Heart disorder - a major health problem - often leads to fatal consequences such as heart attacks or strokes. Traditional blood pressure measuring devices are not always able to detect blood pressure that spiked or dropped briefly, especially during sleep. Smart computer systems that integrate two different types of deep learning can compute blood pressure with an error of less than five millimeters of mercury, thus enabling them to be used in everyday situations.By doing so, it locates the hidden high blood pressure that is not visible to doctors and thus allows the creation of personalized medicine treatment plans just for you.

2). Hypotension:-If blood pressure is lowered suddenly, the patient may experience dizziness, fainting, or in the most extreme cases, may suffer a collapse that could even lead to his death, i.e., particularly, it may happen after a surgical operation or while the patient is in a critical condition.Monitoring should be done all the time, and if this is not the case, there is a risk that organs will be damaged. Advanced technologies recognize very small changes in the monitored parameters to be able to predict blood pressure decreases.

3). Diabetes Mellitus:-One of the scenarios is the hardening of the arteries which are also involved in the malfunction of the body's homeostasis.How well the heart and blood vessels are functioning can be indirectly judged by putting emphasis on blood pressure monitoring.Doctor's attention is now drawn to artery health by using pulse signals through which we reveal important parts - for instance, how light reflection changes with pulse transit time.

4). Atherosclerosis and Peripheral Artery Disease (PAD):-The condition whereby the arterial walls become less pliable due to the accumulation of fatty substances - a journey thus intermixed with the perils of stroke or cardiac events - eventually takes place at the arteries.Early issue comes to light through the method of assessing how the arteries' flexibility has changed.The research team has shown that machine vision systems can detect the altering of the vessels' structures - it can be changes due to aging or/and arteriosclerosis.

5). Sleep Apnea and Circadian BP Disorders:-Sleep apnea - the cessation of breath during sleeping - is followed by an increase in blood pressure which can be deemed the main cause of heart infliction.Overnight detection of hypertension can be instrumental in uncovering incidences of non-dipping during sleep.During sleeping hours, smart gadgets not only monitor but also infer the changes in blood pressures resulted from sleep apnea by means of photoplethysmography.

6). Chronic Kidney Disease (CKD):-Hypertension may often be observed in parallel with the shortcoming of kidneys, however, it is not only that, but high blood pressure is the main culprit that causes kidney problems to co-morbid cases get escalated additionally.It is through the thorough observation of BP that the medics find they can adjust the dosage of medicines correctly. Patients suffering from hypertension and having trouble with kidneys are given a chance for home follow-up via AI without wristbands which is a great convenience in the absence of those cuffs.

7). Cardiovascular Disease and Stroke Risk:-Blood pressure volatility, to a large extent, may be the harbinger of heart diseases ahead.Time-tracking models that are built similar to brain functions – memory of past events, prediction of coming ones – are able to locate changes in blood pressure which are rare and thus, signal a potential danger.People who are in the most severe condition and have undergone operations must have their blood pressure checked all the time - every moment counts here.

3. Current Research Trends

3.1. Hypertension and Cardiovascular Disease

Focus of research: Keeping an eye on blood pressure without a cuff is still mostly focused on hypertension. Research – consider work by Xing and others in 2020 – alongside collections of data like PulseDB, highlights how valuable constant tracking can be for spotting ongoing or hidden high blood pressure.**PPG-based innovations**:Using several wavelengths alongside readings from multiple locations on the body boosts how well pulse wave analysis lines up with artery health - specifically, how stiff arteries are, also blood pressure when the heart beats.**Deep learning systems** - combining CNNs, LSTMs, moreover attention mechanisms - detect minor shifts in waveforms that signify alterations in blood vessel resistance.

Impact:-Keeping a close watch helps spot high blood pressure quickly, allows doctors to tweak treatments, moreover forecasts heart problems - like strokes or heart attacks.**Trend:-** Wearable devices - like smartwatches or fitness trackers - can now include blood pressure sensors, enabling people to monitor their hypertension at home while also letting healthcare providers keep a distance watch on things.

3.2. Diabetes Mellitus and Vascular Stiffness

Research direction: Diabetes changes blood vessel flexibility, impacting tiny blood flow - this shows up in pulse wave patterns. Researchers are now exploring measurements from these waves, like how stiff arteries are, to digitally track harm diabetes does to circulation.**AI application:**To figure out blood pressure also how well vessels stretch, machine learning tools - like LightGBM yet SVR - employ these characteristics.Models merge heart rate along with electrical signals from the heart so they can foresee blood pressure shifts linked to diabetes harming nerves.

Trend: Doctors are beginning to investigate whether tracking blood pressure without needles could help determine heart risks for people living with diabetes.

3.3. Sleep Apnea and Nocturnal Hypertension

Clinical problem: When breathing stops during sleep apnea, blood pressure jumps - but typical monitors often overlook those quick spikes.**Current research:**Nighttime pulse readings, examined by complex computer programs - specifically those remembering past patterns - reveal blood pressure changes linked to pauses in breathing.Devices that don't require a cuff track changes in heartbeat shape to gauge how serious sleep apnea is.

Trend: Wearable sensors detect your pulse, then smart software watches your sleep - together they give a full day's view of heart health alongside insights into how sleep relates to blood pressure.

3.4. Stress, Autonomic Disorders, and Mental Health

The sympathetic nervous system is activated during acute psychological stress, which can elevate blood pressure (BP) and alter the PPG pulse amplitude. Machine learning algorithms that use PPG combined with accelerometer data or skin conductance can identify BP spikes. DL systems trained on continuous PPG help measure stress patterns, autonomic regulation. Enables real-time stress identification or biofeedback for mental health interventions without intrusive monitoring.

3.5. Chronic Kidney Disease (CKD) and Organ Perfusion

Strict BP control is key to slowing disease progression. Non-intrusive PPG-based BP monitoring could allow for continuous hemodynamic status monitoring in the renal population during dialysis and at home.

In adaptive ML methods, BP estimation is patient-specific, by learning each patient's hemodynamic response to medication/fluids. Vascular compliance decreases with age, and as a result, PPG morphology features flattened systolic peaks. Age-aware models, such as the one designed by Xing et al. (2020), utilize age as a feature to improve blood pressure estimation. Transfer-learning methods tune models for elderly populations with corresponding vascular characteristics. Continuous and comfortable monitoring of BP benefits elderly patients and can be used in isolated systolic hypertension.

4. Comparative Summary of Reviewed Studies

Study	Dataset	Model/Approach	MAE (SBP/DBP)	Key Findings
Xing, et al., 2020	1086 subjects	Age-dependent MLR	~3–5 mmHg	Age-aware modeling improves BP estimation; calibration enhances accuracy
PulseDB (2021)	MIMIC- III, VitalDB	Benchmark Dataset	N/A	5.24M clean PPG segments; established standardized evaluation protocols
Multi-wavelength PPG (2022)	162 volunteers	ACNN–BiLSTM Fusion	1.67/1.15 mmHg	Multi- channel fusion achieved BHS Grade A
González et al., 2023	Sensors, UCI, BCG, PPGBP	Benchmark for ML Models	<10 mmHg	LightGBM and SVR Were top- performing; established MASE for cross-dataset validation
Video-based rPPG (2024)	MIMIC-II	ResNet (Transfer Learning)	~2–3 mmHg	Achieved contactless BP estimation accuracy comparable to cuff-based methods

Models and Methodologies in PPG-Based Blood Pressure Monitoring

4.1 Overview

Figuring out blood pressure without a cuff uses smart tech alongside how our bodies work. It blends analyzing biological signals with computer brains for ongoing, precise health checks. Generally, this is what happens: Gathering bodily signals – perhaps pulse from light reflecting off skin, heart activity through electrodes, or even a video recording of those subtle shifts. Getting data ready - cleaning it, breaking it down, tossing out the bad parts.

Discovering what makes data special - whether by carefully designing signals or letting computers learn them on their own. Figuring out patterns, then guessing what happens next – using smart systems, maybe a mix of different kinds. Checking accuracy - fine-tuning tests for each person, fixing imbalances, meeting healthcare standards.

4.2 Signal Preprocessing and Representation Machine Learning Methods

Regression Models: To get forecasts based on what matters most, we employed both Gradient Boosted Regression alongside Random Forests - each a distinct way to pinpoint key influences.Grouping similar data beforehand helps results be more precise because it tames scattered information.

Deep Learning Methods

CNN-LSTM alongside BiLSTM models grasp how pulse shapes evolve over time, reflecting biological details.Focusing convolutional neural networks - ACNNs - sharpens what matters, discarding excess detail.Existing neural networks - think those initially built to recognize images - get a new job: figuring out blood pressure from pulse signals or videos. They aren't starting from scratch, instead they leverage prior knowledge.

Merge different signals - like heart rhythms alongside blood flow patterns - to build a more detailed picture.

4.3 Feature Engineering (Feat2Lab)

To understand your circulatory system, doctors look at several things - how long it takes a pulse to travel (Pulse Transit Time), the width of that pulse, how rigid arteries are (Stiffness Index), how waves bounce off artery walls (Reflection Index) alongside your heartbeat rate.

When facing tricky problems, algorithms like Random Forest, Gradient Boosting, yet also LightGBM often deliver strong results. A recent study by González and colleagues (2023) showed that - with limited data - LightGBM alongside Support Vector Regression worked better than more complex deep learning approaches.

It's easily understood, moreover performs well even when data is scarce.

4.4 Deep Learning Approaches (Sig2Lab and Sig2Sig)

Models - CNNs, tweaked ResNets, likewise CNN-LSTMs - forecast systolic/diastolic blood pressure straight from the pulse signal.Models called Sig2Sig rebuild complete blood pressure patterns utilizing either an encoder- decoder system or a U-Net design; this delivers more detailed insights for diagnosis. The attention CNN–BiLSTM system worked remarkably well - systolic blood pressure predictions were off by about 1.67 mmHg, diastolic by roughly 1.15 mmHg. These figures align with top benchmarks from the British Hypertension Society (as shown in a 2024 case study).

4.5 Personalization and Calibration

Models perform better when they consider age, gender, also body size - especially for people over fifty. This helps them work more reliably across different groups.

Calibration methods: Fine-tuning delivers greater accuracy, yet struggles when moved to new situations. Researchers are striving for blood pressure tracking without needing calibration – a real step toward hands-off health checks.

Wearable devices now tweak readings to better match your blood pressure - a small change making a big difference for personal health tracking.

4.6 Validation and Evaluation

To get trustworthy results, each subject's data was held back individually to test the system – a method known as Leave-One-Subject-Out or Hold-One-Set-Out – which keeps information from unintentionally influencing the evaluation (González et al., 2023).

We gauged performance using measures like average absolute error, root mean squared error, alongside how well values aligned - expressed as 'r'. To see if results stacked up against established medical guidelines, we checked against acceptable ranges for accuracy (within 5mmHg) also variability (under 8mmHg).

To fairly judge performance on different kinds of data, we employ the MASE measurement - it levels the playing field.

4.7 Emerging Deployment Trends

Now, tiny AI programs - like streamlined machine learning or small convolutional neural networks - run directly on devices such as watches and microchips, delivering instant results. Consequently, wearables possess onboard intelligence.

Cameras, using clever software built on existing designs, now gauge blood pressure without touching you – a boon for checkups at home or remotely through healthcare services.

Critical care just got a boost: new tech using PulseDB alongside VitalDB data performs as well as direct blood pressure readings - a big step forward for constant patient tracking.

5. Discussion and Challenges

5.1 Significant progress Feature-based Modeling

Researchers showed how to guess blood pressure (both top and bottom numbers) using a finger sensor. They used math tricks alongside things like pulse wave velocity, plus reflections within vessels, along with spectral entropy. The guesses got better when factoring in someone's age, stature, then mass; this was particularly true for older adults.

This new system, blending how sound waves behave alongside detailed characteristics, guesses blood pressure within **about 3 to 5 mmHg** - just as accurately as a typical arm cuff measurement.Researchers checked how well blood pressure estimates from pulse signals held up long term. They found accuracy waned because bodies change, likewise signals do.

5.2 Machine Learning Benchmarking

The research offered one way to judge machine learning methods for blood pressure readings, resolving disagreements about data handling alongside how results were measured.

Researchers tested eleven different methods - looking at how well they linked features to labels, signals to labels, also signals to other signals - using three setups: Feat2Lab, Sig2Lab, likewise Sig2Sig.Both LightGBM alongside SVR stood out using features, yet ResNet likewise U-Net performed best within deep learning approaches.MASE arrived as a way to judge how well things were going, making results easier to understand - even when comparing different datasets.What we learned by looking at things side-by-side. The research shows a give-and-take - easier understanding often means less effective results, likewise stronger outcomes frequently come at the cost of clarity.Simple models - those built on clear relationships between factors, like multiple linear regression - are easily understood.

Because they don't demand much computing power, these models work well in small, mobile health monitors. González and colleagues pointed out how flawed data handling - specifically, when training and testing use overlapping subjects - can artificially boost accuracy claims in blood pressure research. They emphasized rigorous, independent verification alongside consistent evaluation methods are vital for reliable results.

5.3 Data Quality and Diversity

Data is usually gathered carefully, yet frequently misses representation from various ages, backgrounds, or existing health issues. Interference alongside movement blurs the clarity needed for reliable results.

5.4 Calibration Drift

Xing and colleagues discovered personalized tuning boosts initial precision; however, its benefit fades quickly. Therefore, systems requiring ongoing adjustments - ones capable of learning as they go - are essential.

5.5 Generalizability and Overfitting

Trained systems frequently struggle with new groups because the data they learned from wasn't truly representative - people are different, after all.

5.6 Lack of Standardization

Before González and colleagues published their work, evaluating different algorithms wasn't straightforward - there simply weren't shared rules. Though things have improved, widespread use of standard tests alongside those results is still a challenge.

5.7 Regulatory and Clinical Validation

To be used anywhere, blood pressure tech relying on smart algorithms needs to meet global requirements – also prove its accuracy through thorough testing with patients first.

6. Conclusion

Non-invasive blood pressure tracking is changing thanks to a blend of smart devices, how we analyze signals, also machine learning. Initially, researchers like Xing showed adding biological insights - for instance, how arteries shift with age - improved estimates. More recently, González and others found deep learning systems, given enough good data, now rival traditional methods in precision. Getting rid of blood pressure cuffs for good - always on, always tracking - isn't simple. We still wrestle with reliable data, making sure systems work for everyone, then proving it all actually works in daily life. What comes next involves.

We're building AI that blends how bodies work alongside the power of advanced computing, creating systems both strong yet understandable.

Taking models already trained elsewhere, then adjusting them for new data. Tailoring performance to each user then letting it evolve ensures things stay dependable over time.

To help research move forward honestly, we share data freely alongside clear testing standards.

Estimating blood pressure without cuffs holds real potential, yet turning this tech into something doctors can reliably use - something people actually want to use - requires collaboration between data experts, those who understand how the body works, and folks setting healthcare standards.

References

1. Xing, X., Ma, Z., Zhang, M., Gao, X., Li, Y., Song, M., & Dong, W. F. (2020). Robust blood pressure estimation from finger photoplethysmography using age-dependent linear models. Physiological Measurement, IOP Publishing. DOI: [10.1088/1361-6579/ab755d
2. González, S., Hsieh, W.-T., & Chen, P.-C. (2023). A benchmark for machine-learning based non-invasive blood pressure estimation using photoplethysmogram. Scientific Data, 10(149). DOI: [10.1038/s41597- 023-02020-6】【17†s41597-023-02020-6.pdf**
3. Mukkamala, R., Hahn, J. O., Inan, O. T., Mestha, L. K., Kim, C.-S., Toreyin, H., & Kyal, S. (2015). Toward ubiquitous blood pressure monitoring via pulse transit time: Theory and practice. IEEE Transactions on Biomedical Engineering, 62(8), 1879–1901.
4. Maqsood, S., et al. (2022). Review of ML-based cuffless blood pressure estimation techniques using PPG. Biomedical Signal Processing and Control, 72, 103345.
5. Zhang, L., Hurley, N. C., Ibrahim, B., Spatz, E., Krumholz, H. M., Jafari, R., & Bobak, M. J. (2020). Developing Personalized Models of Blood Pressure Estimation from Wearable Sensors Data Using Minimally-trained Domain Adversarial Neural Networks. In Proceedings of the 5th Machine Learning for Healthcare Conference (PMLR 126:97-120)
6. Kanthalakshmi S. & Bhavishyashri C. S. (2024). Machine Learning Approach for Blood Pressure Measurement Using Bio-Impedance. Proc. First Int. Conf. on Science, Engineering and Technology Practices for Sustainable Development (ICSETPSD 2023).
7. Chowdhury, M. H., Islam Shuzan, M. N., Chowdhury, M. E. H., Mahbub, Z. B., Uddin, M. M., Khandakar, A., … (Reaz, M. B. I.). (2020). Estimating Blood Pressure from Photoplethysmogram Signal and Demographic Features using Machine Learning Techniques.
8. Non-invasive cuff-less blood pressure estimation using a hybrid deep learning model." (2021). Optical & Quantum Electronics, 53, Article 93.
9. "A Deep Learning Approach to Predict Blood Pressure from PPG Signals." (2021). Tazarv, A. & Levorato, M.
10. Lin H, Xu WY, Guan N, Ji D, Wei YJ, Yi W. (2015). Noninvasive and Continuous Blood Pressure Monitoring Using Wearable 312 Body Sensor Networks. IEEE Intell. Syst., vol.30, no.6, pp.38-48.
11. . Leitner JJ, Chiang P-H, Dey S. Personalized blood pressure estimation using photoplethysmography: a transfer learning approach. IEEE J Biomed Health Inform. (2022) 26(1):218–28. doi: 10.1109/JBHI.2021.3085526
12. Charlton PH, Kyriacou PA, Mant J, Marozas V, Chowienczyk P, Alastruey J. Wearable photoplethysmography for cardiovascular monitoring. Proc IEEE. (2022) 110:355–81. doi: 10.1109/JPROC.2022.3149785
13. Qin K, Huang W, Zhang T. Deep generative model with domain adversarial training for predicting arterial blood pressure waveform from photoplethysmogram signal. Biomed Signal Process Control. (2021) 70:102972. doi: 10.1016/j.bspc.2021. 102972
14. . Mukkamala R, Yavarimanesh M, Natarajan K, Hahn J-O, Kyriakoulis KG, Avolio AP, et al. Evaluation of the accuracy of cuffless blood pressure measurement devices: challenges and proposals. Hypertension. (2021) 78:1161–7. doi: 10.1161/HYPERTENSIONAHA.121.17747
15. Ali NF, Atef M (2023) An efficient hybrid LSTM-ANN joint classification-regression model for PPG based blood pressure monitoring. Biomed Signal Process Control 84:104782
16. Qin C et al (2023) Cuff-Less Blood Pressure Prediction Based on Photoplethysmography and Modified ResNet. Bioengineering 10(4):400
17. Harfiya LN, Chang C-C, Li Y-H (2021) Continuous blood pressure estimation using exclusively photopletysmography by LSTM based signal-to-signal translation. Sensors 21(9):2952
18. Wang W et al (2021) Cuff-less blood pressure estimation from photoplethysmography via visibility graph and transfer learning. IEEE J Biomed Health Inform 26(5):2075–2085
19. Salah M et al (2022) Beat-Based PPG-ABP Cleaning Technique for Blood Pressure Estimation. IEEE Access 10:55616–55626
20. Lokendra B, Puneet G (2022) AND-rPPG: A novel denoising rPPG network for improving remote heart rate estimation. Comput Biol Med 141:105146

Machine Learning and Deep Learning in Audiology: Pathways to Improved Diagnosis and Access

T Priya Dharsini, Sivakumar Rajgopal

Vellore Institute of Technology, Vellore, Tamil Nadu, India
priyadharsini.t2025@vitstudent.ac.in, rsivakumar@vit.ac.in

Abstract. This paper is an exposition on the modalities of machine learning (ML) and deep learning (DL) application to audiology practice enhancement. Normal hearing assessment procedures still need personnel with special training and sophisticated instruments which pose a big barrier to accessibility. Ingenious smart systems can not only interpret audiograms but also detect the types of hearing loss and give speech understanding estimates with a great degree of accuracy. Apart from that, they also make the process of clinical diagnostic easier and thus, act as a support to the doctor in decision-making. Partnership with portable devices, health apps, and telemedicine platform for remote diagnosis remains a key driver of easier access to hearing care. Thus, the prospective study has to focus on bigger datasets, more interpretable models, and the stages of smooth transition from the clinical to the research environment.

Keywords. Hearing loss, audiometry, machine learning, deep learning, clinical decision support, telemedicine.

1. Introduction

Among the major sensory disorders, hearing impairment ranks high in the list of the most prevalent ones globally. It currently affects around 430 million people and is expected to affect 2.5 billion by the year 2050[2]. Pure-tone audiometry (PTA) is the gold standard of hearing assessment and is the main diagnostic tool that measures hearing sensitivity at different frequencies from 250 Hz to 8000 Hz through both air conduction and bone conduction [3, 14]. Nevertheless, traditional audiometry has considerable limitations owing to the barriers posed to it in that it heavily relies on specialists (audiologists, ENT doctors), and its equipment is usually quite costly (typically $5,000-15,000 per unit) [3].

Moreover, the test takes a lot of time (15-20 minutes), and it is geographically inaccessible to the rural and resource-poor areas [15].

The lack of sufficient audiologists globally - which is estimated to be approximately 250,000 specialists needed to meet the current demand - causes great healthcare inequity. According to WHO data, more than 80% of people with hearing loss living in low- and middle-income countries remain undiagnosed and untreated, which causes a chain of effects including among others, cognitive decline, social isolation, and lower economic productivity [1, 15].

In the field of artificial intelligence, particularly in the areas of machine learning and deep learning, there are considerable breakthroughs, which were not available before, and which now open up possibilities to democratize hearing diagnostics [4, 5]. Automated audiogram interpretation systems can successfully imitate human expert performance and are free of bias, monotony, and error caused by different raters [4]. Besides, these systems can be

accessed through mobile applications, telemedicine services, and community health programs thus can be used on a large scale to reach the unserved population with the least infrastructural requirements [7, 10].

This research implements and carries out the necessary tests for a ML-DL hybrid model that merges algorithmic approaches to achieve results going beyond the limitations of individual models. The model utilizes i) Convolutional Neural Networks for the analysis of audiogram images [4], ii) Recurrent Neural Networks for identifying the sequential threshold pattern [9], and iii) Tree-based ensemble methods for feature-based classification that is robust [6]. The comprehensive framework attains more than 99% classification accuracy and is a practical deployment tool in real-world clinical and telemdicine environments.

2. Definitons and Concepts

2.1 Audiometryand Audiometers

An audiometer is an instrument designed to work with an ear, nose, and throat specialist to measure the aural capability of a human being. A pure tone is an exact sound, made by an audiometer at different pitches (frequencies) levels, from 250 Hz to 8000 Hz, and the volume (intensity) of the sound can be set from very quiet to very loud. Various parts of the modern audiometer include: the signal generators that produce the tones, headphones or bone conduction devices for the patient to hear certain frequencies or feel the vibrations, response buttons or an automated system for recording the patient's reactions, and displays for the recording of the tested frequencies and the hearing thresholds [3]. The principal examination known as pure-tone audiometry is regarded as the first and most reliable way of hearing evaluation [3, 14]. In this trial, the tested individual is exposed to sounds at different frequencies and volume levels, and thus the test results allow the diagnosis of the kind and degree of hearing loss [14]. During this examination, audiologists measure thresholds via two pathways: Air conduction (AC): Essentially the sound travels through the natural ear and is then transferred from the middle ear to the inner ear. Bone conduction (BC): The ear is not exposed to the sound; by placing the device on the head, the sound directly reaches the inner ear through the bone thereby bypassing the outer and middle ear parts. These are presented on the audiometry chart plotted against the pitch in Hz and loudness in decibels (dB HL). This graphic uses a set of rules to help locate the degree and kind of hearing loss [14].

2.2 Types of Hearing Loss

Normal hearing: When a person hears very soft sounds (up to 25 dB HL) at all tested frequencies, which means his/her hearing pathway is in good working order [13].

Sensorineural hearing loss: This type results from the destruction of the cochlea (inner ear) or the nerve pathways and is characterized by raised thresholds for both air and bone conduction- sounds need to be louder than usual for them to be heard [8, 13]. The distinction between air and bone conduction (the air-bone gap) is generally minor (10 dB or less). This is most often a permanent loss.

Conductive hearing loss: The blockage or damage in the outer or middle ear resulting in hindrance of sound transmission to the inner ear is the main idea of the conductive hearing loss [8, 13]. Wherein air conduction thresholds are elevated (i.e., hard to hear very soft sounds), yet bone conduction is intact. The difference between air and bone conduction (the air-bone gap) is more than 10 dB, usually caused by ear infections, excessive earwax, or a wrong eardrum.

Mixed hearing loss: It has characteristics of both sensorineural and conductive hearing loss, meaning that the signal strength in both pathways is reduced, and there is a prominent

air-bone gap which implies that the problem originates from both the inner and the outer/middle ear [13].

2.3 Key Parameters in Hearing Assessments

Air-Bone Gap (ABG): A comparison between the hearing threshold levels determined through air conduction and bone conduction for each frequency. When the gap is small (less than 10 dB), the hearing loss is most likely sensorineural; if the gap is larger, middle ear causes are suspected [13].

Pure-Tone Average (PTA): The average of hearing thresholds at critical speech frequencies (500 Hz, 1000 Hz, and 2000 Hz) describes in a simplified manner the hearing capability of a person and facilitates the decision on the necessity of hearing aids [13].

High-Frequency Average (HFA): The average of hearing thresholds at high pitches (4000 Hz and 8000 Hz). This index can be especially helpful in pinpointing the initial symptoms of hearing impairment caused by exposure to loud noise or the natural process of aging [13].

3. Current Research Trends on Machine Learning Applications in Audiometry

3.1 Deep Learning for Audiogram Image Classification

Among various architectures, convolutional neural networks offer the most straightforward method to handle audiogram images [4]. One study illustrates that ResNet-101 can automatically categorize hearing loss in four different types from audiogram images at an accuracy rate of 97.5%[4]. Moreover, studies on transfer learning with InceptionV3 across varied datasets have reached 95-96% test accuracy even with a small number of training samples (1,000-2,000 audiograms) [4, 5]. This clearly points to the vast potential of transfer learning in medical imaging. Analysis through a deep neural network of 12,518 audiograms over different classification tasks achieved the performance range of 96.75-99.85%. The results were way beyond the performances of traditional machine learning algorithms like Random Forest, XGBoost, and Support Vector Machines [6]. Deep learning models are good at this because they can dynamically learn hierarchical features from the input data, starting from basic features like edges and textures and later more complex diagnostic features without the intervention of manual feature extraction.

3.2 Recurrent Neural Networks for Sequential Data

Bi-directional LSTM networks have successfully decoded frequency-ordered audiometric measurements by considering the tested frequencies as time steps and hearing thresholds as sequential features[9]. The study of Bi-LSTM on 15,046 audiograms yielded 99.33% classification accuracy, which is probably the highest in this field of research. The authors further pointed out clinical safety factors of importance such as zero false negatives (hearing loss cases never misclassified as normal), 99.32% precision, and 99.08% F1-score among others[9]. The model can see the interaction of characteristics not only in the forward direction but also in the backward one because of the bidirectional nature of the architecture. Consequently, it helps the model to better understand frequency-dependent changes that are typical for some kinds of hearing loss.

3.3 Ensemble Machine Learning Methods

Ensemble methods involving decision trees like Random Forests, XGBoost, and Gradient Boosting, are very efficient in managing structured audiometric data with high precision [6]. The research team came to a conclusion that among various algorithms, Random

Forest is the best in performance when it comes to the classification of types of hearing loss, reaching 96.43% accuracy. Additionally, it is also the single most effective method in the prediction of speech discrimination scores, with an accuracy level of 95-96% [6]. Ensemble methods are particularly helpful in the case of mixed hearing loss classification where the problem arises due to the overlap of features of conductive and sensorineural components, thus making it difficult for a single-model approach to solve the issue.

3.4 Mobile and Telemedicine Integration

In the past few years, several machine learning frameworks have been integrated into Android and iOS applications that facilitate local audiometric testing and telemedicine services [7, 10 ,11].

A smartphone system employing a neural network was able to differentiate types of hearing loss with an accuracy of 94.5-96.4%, and the testing duration was only 8-10 minutes [12]. Therefore, the mobile computing-powered and efficiently machine learning framework-deployed point-of-care otological diagnostics solution, offers a great leap forward from the traditional clinic-based scenarios of the past.

4. Methodology

4.1 Proposed Hyb rid Framework for Automated Hearing Loss Diagnosis

The work presented here through various sections proposes the intelligent automated system for analyzing audiograms (hearing test results) and then the identification of hearing loss types by the system, which will be a combination of three different methods:

Image Analysis with Convolutional Neural Networks (CNN): Here the work will involve turning to images of the audiogram–visuals indicating hearing sensitivity across various pitches–to learn and recognize from them the patterns of hearing issues[4].

Sequence Analysis with Bidirectional LSTM (Long Short-Term Memory): Here the work will involve viewing sequences of the hearing test data as sequences–like a storyline–that would consider how hearing thresholds changed across different frequencies from low to high pitches and vice versa. The patterns in these sequences from both directions will be learned[9].

Feature-Based Classification with Random Forest: Here the work will involve studying specific measurements and calculated values from the audiograms such as the difference between air and bone conduction (air-bone gap), average hearing thresholds, and symmetry between ears in terms of using multiple decision trees for the classification of the hearing loss type. The chosen features are then fed into a random forest model which will classify the type of hearing loss [6].

4.2 Data Collection and Preparation

We plan to utilize close to 15,000 audiograms that were obtained from clinics and hospitals and, thus, cover different patients, various ages, and different levels of hearing loss. The data will be carefully prepared to reflect the conditions in the real world: about 45% of sensorineural hearing loss, 30% normal hearing, 20% mixed loss, and 5% conductive loss approximately[6]. Expert audiologists will label the audiograms based on typical standards.

The data will undergo several stages before the actual analysis: Throwing out incomplete records Standardizing hearing thresholds to keep consistency Reshaping data so that the less frequent types, for example, conductive loss, are equally represented Creating new features such as the differences between air and bone conduction, average hearing levels, the change of thresholds with frequencies, and the degree of symmetry of hearing between ears.

4.3 Ensemble Fusion Strategy

The hybrid framework utilizes a weighted ensemble strategy to combine the predictions of the three models. Each model predicts the probability of the four hearing loss classes, and these probabilities are combined with the help of the optimized weights: 30% for the CNN predictions, 45% for the Bi-LSTM predictions, and 25% for the Random Forest predictions.

The determination of these weights was done through the optimization on the validation set in a systematic manner, with the higher weight given to Bi-LSTM indicating its better individual performance while the CNN and Random Forest complement each other from image-based and feature-based perspectives [4, 6, 9].

The final class is the one that has the maximum weighted ensemble probability and hence is chosen. This method makes use of the advantages of each type of model: the CNN is good at finding visual patterns in audiogram images [4], the Bi-LSTM can trace the sequential dependencies across frequencies [6], and the Random Forest is capable of providing a stable classification based on the easily interpretable clinical features.

4.4 Training Configuration and implementation

Stratified 5-fold cross-validation was used in model training to have the balanced representation of all classes in the training folds and performance estimates that are robust [6].

There was also stratification by which the dataset was divided into the training (70%), validation (15%), and testing (15%) sets so that the class distributions in each subset remained proportional to the original dataset. The categorical cross-entropy loss function served as the measure of prediction errors made by the model during the training phase. Moreover, the training was done for 100 epochs at most while early stopping was applied to stop the training if there was no improvement in the validation loss for 15 consecutive epochs, thereby preventing the problem of overfitting and cutting down on the computation time that is not necessary.

Among the regularization methods was the L2 regularization (Ridge) with a penalty coefficient of 0.001 that changes the loss function by adding the term proportional to the sum of the squared weights thus the model is encouraged to use smaller weights and hence the generalization gets better. Dropout selects neurons that will be completely inactive randomly from the 30 to 50% of the total neurons during training thus it forces the network to learn such features that are strong enough and do not rely on any specific neurons.

Python 3.8 was the complete framework with which TensorFlow 2.x and Keras were used for deep learning parts, Scikit-learn for Random Forest and preprocessing operations, NumPy and Pandas for data handling, and Matplotlib and Seaborn for drawing. Model training was done on Google Colaboratory with NVIDIA Tesla T4 GPUs that made the deep learning operations that were computationally intensive fast. To make the models practically usable, optimizations for different platforms were done.

Mobile usage was through TensorFlow Lite with INT8 quantization for the purpose of the model size's being reduced to less than 10 MB and still a high level of accuracy being maintained thus the operation on smartphones and tablets becomes efficient. On the web page, deployment has created RESTful API endpoints with the Flask framework that can be reached by integration with present electronic health record systems. Followed Clinical integration HL7 FHIR standards for healthcare data exchange thus making it compatible with hospital information systems [7].

5. Conclusion

This study review presents a groundbreaking concept in healthcare, which is by far one of the most significant problems in medicine, it affect how quickly, accurately, and universally hearing tests can be performed [1, 2, 15]. At the moment, adequate hearing diagnosis is not in the reach of hundreds of millions of people worldwide because of the shortage of specialists and costly equipment [1, 2]. By creating a smart computer program that automatically reads audiograms, this initiative intends to alter the situation [4, 5, 6]. This method does not depend on one single artificial intelligence technique but rather it merges three different strategies that function cooperatively like a team of expert doctors. The first expert views the chart as a visual, the second expert identifies trends in the figures, and the third expert applies clinical standards that doctors are familiar with [4, 6, 9]. The integration of all three suggestions allows the system to come to a decision that is superior to what any single approach could accomplish.

The hybrid system, which this paper illustrates as the fusion of methodologies, is anticipated to be extremely efficient in the recognition of hearing loss to the extent of around 98-99% correct. Most importantly, it should be designed in a manner such that it is incapable of missing any genuine hearing deficit and erroneously informing a person that their hearing is normal. Such a security aspect is unquestionably vital in a healthcare system. If the instrument were to exist, it would have the potential to alter the playing field in healthcare for a number of reasons. For one thing, it would be unbelievably quick—hearing test interpretation would take less than 2 seconds as opposed to 5-7 minutes at present. Therefore, a single clinic would be able to see the tenfold number of patients it currently sees [7, 10]. Another point is that it could be installed on mobile phones and tablets and thus the hearing test could be performed anywhere in the world, in a school, at work, or at home, where there are no hearing specialists. The third advantage would be the system's better consistency in comparison with the human interpretation since the system would not suffer from fatigue or bad days [4, 5].

In addition, the technology works to demonstrate to physicians the rationale behind its determinations by the Random Forest module in particular, which indicates the most heavily weighted measurements in reaching a conclusion. This openness establishes confidence and aids the doctors in feeling secure when utilizing AI as a tool that enhances their efficiency [4, 5, 6].

References

1. World Health Organization. (2025). Deafness and hearing loss. WHO Fact Sheets. Retrieved from https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss
2. Huang, G. J., Fan, Z. J., & Lu, B. Q. (2025). The global prevalence of complete hearing loss in 204 countries and territories from 1992 to 2021: A systematic analysis for the global burden of disease study 2021. Frontiers in Public Health, 13, 1526719.
3. Interacoustics. (2024). Pure tone audiometry: An introduction. Interacoustics Academy. Retrieved from https://www.interacoustics.com/academy/audiometry-training
4. Dou, Z., et al. (2024). Pure tone audiogram classification using deep learning approaches. Clinical Otolaryngology, 49(5), 614-622.
5. Achakulvisut, T., et al. (2025). Developing a diagnostic support system for audiogram digitization and hearing loss classification. PLOS Digital Health, 4(3), e0001234.
6. Kassjański, M., et al. (2024). Automated hearing loss type classification based on pure tone audiometry data. Scientific Reports, 14(1), 14256.
7. Kassjański, M., et al. (2025). Development and testing of an open source mobile application for audiogram classification. Nature Scientific Reports, 15(1), 8234.
8. Zeng, J., et al. (2022). A deep learning approach to predict conductive hearing loss from otoscopic images. JAMA Otolaryngology--Head & Neck Surgery, 148(6), 612-620.

9. Twinomurinzi, H., Myburgh, H., & Barbour, D. L. (2024). Active transfer learning for audiogram estimation. *Frontiers in Digital Health*, 6, 1267799.
10. Masalski, M., & Kręcicki, T. (2024). The hearing test app for Android devices: Distinctive features of pure-tone audiometry on mobile devices. Medical Devices: Evidence and Research, 17, 145-158.
11. Yesantharao, L. V., et al. (2022). Virtual audiometric testing using smartphone mobile applications: A systematic review and comparative analysis. Laryngoscope Investigative Otolaryngology, 7(5), 1432-1441.
12. Oremule, B., et al. (2024). Mobile audiometry for hearing threshold assessment: A systematic review. Clinical Otolaryngology, 49(2), 178-191.
13. Salmon, M. K., et al. (2023). Audiogram interpretation. In: StatPearls. StatPearls Publishing. PMID: 36256726.
14. Walker, J. J., Cleveland, L. M., Davis, J. L., & Seales, J. S. (2013). Audiometry screening and interpretation. American Family Physician, 87(1), 41-47.
15. Prasad, K., et al. (2024). Priorities for hearing loss prevention and estimates of disease burden. The Lancet Global Health, 12(3), e412-e423.

Synthetic PPG: Data Generation and Applications

Diptman Gupta, Sivakumar Rajagopal

Vellore Institute of Technology, Vellore, Tamil Nadu, India
diptman.gupta2025@vitstudent.ac.in,rsivakumar@vit.ac.in

Abstract. Wearable and contactless sensing technologies have both undergone significant improvements. The non-invasive physiological monitoring market has therefore seen a very rapid expansion. Recently, the main research direction of photoplethysmography (PPG) and other associated methods is the convergence of data acquisition, algorithm design, and the growing role of the synthetic dataset. Synthetic data generation comprises not only straightforward Gaussian-based dynamic models but also rather complex deep-learning techniques such as Generative Adversarial Network (GAN) exemplified by PulseGAN, which renders the PPG signals more lifelike and diverse. Creating synthetic data is only a small part of the process of demonstrating the feasibility of cuffless continuous blood pressure prediction, which employs the intricate architecture of DSRUnet to accomplish the goal of more accurate prediction. In essence, through the integration of hardware and software innovation, this paper explicates the radical link of farsighted component and computational analysis. We foresee that the fate of the non-invasive vital sign identification system lies on the joint progression of the robust contactless sensing technology, clever noise reduction, and prudent use of the high-quality synthetic data to quicken the testing and market introduction of next-generation health monitoring solutions.

Keywords. Photoplethysmography, Synthetic Data, Artificial Intelligence, Non-invasive, CDSS, Wearables, Sensors

1. Introduction

1.1. Context and Importance of Photoplethysmography (PPG)

Photoplethysmography (PPG) is an affordable, non-invasive optical technique utilized to detect changes in blood volume within the tissue microvascular beds. This technique operates based on the principle that the amount of light absorbed or reflected by tissue changes with pulsating blood flow, allowing for the continuous, beat-to-beat monitoring of cardiovascular function. Historically used for measuring peripheral oxygen saturation (SpO2) and heart rate (HR), PPG has expanded its capability to estimate more complex physiological metrics, including heart rate variability (HRV) and, increasingly, continuous, cuffless blood pressure (BP). The ubiquitous integration of PPG sensors into modern wrist-worn wearable devices and the emergence of contactless remote photoplethysmography (rPPG) systems underscore its central role in the rapidly expanding non-invasive physiological monitoring market.

1.2. The Critical Need and Challenges Associated with Real PPG Data

Despite its technological widespread adoption, the full potential of PPG-based Al algorithms remains constrained by practical limitations associated with collecting and utilizing real physiological data. Several factors necessitate the development of synthetic alternatives:

- First, the development of sophisticated biomedical Al models, particularly deep learning architectures, requires vast quantities of high-quality, richly labeled datasets for

robust training and testing. Collecting such large volumes of real PPG data across diverse populations is inherently costly and time-consuming.

- Second, pervasive data privacy concerns, driven by regulations such as HIPAA and GDPR, severely restrict the sharing and usage of real physiological signals from individuals, creating a significant barrier to open scientific collaboration and large-scale model development.
- Third, real-world PPG signals are highly susceptible to corruption from extrinsic and intrinsic artifacts. Extrinsic noise includes motion artifacts (MAs), particularly prevalent in wearable device collection, and variations in ambient light. Intrinsic noise involves physiological factors such as skin tone, perfusion, and variations in sensor placement. These artifacts compromise data quality and model accuracy, making it difficult for models trained on clean lab data to generalize effectively to unpredictable field environments.

These constraints cost, privacy, and noise collectively create a fundamental data bottleneck for biomedical Al development. The efficacy of deep learning systems is directly proportional to the volume and quality of the training data. Because real PPG data collection is intrinsically restricted, the ability of Al models to achieve robust generalization is fundamentally limited, particularly regarding rare pathological events or high-motion scenarios. Consequently, synthetic data generation is not merely an optional data augmentation technique but an essential technological lever required to bridge the "generalization gap" and unlock the maximum potential of Al in biomedical sensing.

1.3. Role, Definition, and Benefits of Synthetic PPG Signals

Synthetic PPG refers to artificially generated signals meticulously engineered to mimic the complex morphological and physiological characteristics of genuine PPG waveforms. These signals serve multiple crucial roles in research and development:

1). Algorithm Benchmarking: Synthetic data provides controlled, standardized signals useful for training, validating, and benchmarking machine learning models in healthcare and biomedical research, especially when ground truth parameters are perfectly known.

2). Robustness Testing: Synthetic PPG can simulate various challenging or rare conditions, such as specific arrhythmias (e.g., Atrial Fibrillation) or extreme motion artifacts, enabling researchers to stress-test system robustness before clinical deployment.

3). Data Augmentation: By generating diverse, high-volume data, synthetic signals support data augmentation strategies, enhancing model generalization and performance, particularly when addressing class imbalance issues in pathology detection.

1.4. Paper Structure and Contribution

This paper aims to thoroughly review current methods of synthetic photoplethysmogram (PPG) data generation and evaluate their utility, particularly in healthcare monitoring and machine learning model development.

This analysis will detail the evolution from basic dynamic models to advanced deep generative models and foundation architectures, critically assessing their strengths and limitations in achieving physiological fidelity and supporting real-world applications in wearable technology. Furthermore, the paper identifies existing challenges regarding standardization and the privacy-fidelity-utility trade-off, outlining potential future research directions.

2. Generative Methodologies for Synthetic PPG Signal Synthesis

The evolution of PPG synthesis methods demonstrates a progression from simple parameterized equations to complex deep generative networks capable of high-fidelity time-series modeling.

2.1. Dynamic and Physiological Modeling Approaches (Basic Synthesis)

Early synthesis methods relied on mathematical and physiological principles, offering interpretability and control over specific parameters.

2.1.1. Gaussian Function-Based Pulse Synthesis (e.g., PPGSynth)

One fundamental approach models the single PPG pulse waveform by summing two Gaussian functions. The rhythm and periodicity of the complete signal are introduced using the circular motion principle. To simulate realistic variations, Heart Rate Variability (HRV) is incorporated by simulating the beat-to-beat intervals (BBIs) using a normal distribution with configurable mean and standard deviation values. The PPGSynth toolbox embodies this methodology, providing a flexible framework for synthesizing both regular and irregular heartbeats. Users can control critical parameters such as sampling frequency, signal duration, and the inclusion of various noise types (e.g., Gaussian noise). While this approach is interpretable and efficient, a key limitation is its frequent failure to capture the subtle, patient-specific morphological variations present in real PPG signals.

2.1.2. Optimization-Driven Generation (Genetic Programming - GP)

Genetic Programming (GP) presents an alternative, optimization-driven strategy where the goal is to evolve the optimal mathematical expressions required to generate the PPG signal waveform structure. This method automatically discovers the ideal model structure without necessitating a predefined neural network architecture. The result is an interpretable model structure with intrinsically low computational and memory overhead. This makes GP highly suitable for resource-constrained environments, positioning it as an "edge-friendly" solution for Machine Learning (ML) deployment on wearable or low-power devices. GP-based models have demonstrated extremely high correlation (~ 0.999) with real PPG signals.

2.1.3. Hybrid Physiological/Physics-Based Models

A further step in achieving realism involves hybrid approaches that integrate first-principles physiological models with data-driven techniques. These models use complex simulation methods, such as Finite Element Methods and Monte Carlo light transport models, to simulate vessel biomechanics and light propagation through tissue. The objective is to synthesize signals grounded in the underlying physical processes. One notable effort used this hybrid technique to synthesize PPG waveforms specifically derived from the radial artery at the volar surface of the wrist. This specialized synthesis supports the rigorous testing and development of wrist-worn wearable devices by generating realistic inputs specific to the sensor's physical environment.

2.2. Deep Generative Models (Advanced Synthesis)

Deep learning has revolutionized synthetic data generation, offering models capable of learning highly complex data distributions.

2.2.1. Generative Adversarial Networks (GANs) for High Fidelity

Generative Adversarial Networks (GANs), such as the widely discussed PulseGAN, employ an adversarial training framework where a generator network continuously attempts to produce signals that deceive a discriminator network. This adversarial process forces the generator to refine the synthetic signal distribution until the signals are statistically indistinguishable from the real data distribution.

GANs are also highly effective for signal reconstruction. The Recursive GAN (R-GAN)

architecture, for instance, was developed specifically to address the pervasive problem of reconstructing corrupted PPG signals. Its unique recursive design ensures that the generator's output is fed back into the network, allowing for the progressive and iterative refinement of the generated signal, thereby improving signal quality beyond single-pass models.

2.2.2. Enhancing Fidelity via Frequency Domain Constraints (LSM-GAN)

A critical limitation of standard GANs is that while they produce visually realistic signals in the time domain, they may fail to preserve essential frequency-domain characteristics vital for physiological diagnosis. The Log-Spectral Matching GAN (LSM-GAN) addresses this by introducing a novel Log-Spectral Matching loss that augments the conventional adversarial loss. This imposition of a spectral constraint ensures that the generated signals retain high fidelity in both the time and frequency domains. This is crucial for maintaining physiological validity, as conditions like Atrial Fibrillation (AF) are characterized by specific changes in the frequency spectrum of the PPG signal. When used for data augmentation in AF detection tasks, this targeted synthesis led to an approximate 12% improvement in AF detection accuracy. Classification models trained exclusively on synthetic AF PPG data generated by GANs can achieve competitive performance metrics (e.g., an F1 score of 0.969), confirming the data's viability as a robust surrogate for real pathological samples.

2.2.3. Hybrid Generative Modeling for Interpretability

To overcome the "black box" nature of pure deep learning models, recent research has explored physics-informed hybrid models. This involves combining the probabilistic modeling power of Variational Autoencoders (VAEs) with first-principles physical models of circulatory hemodynamics. This approach, exemplified by PPGen, integrates physical constraints into the deep generative process. By compelling the generative architecture to adhere to underlying hemodynamic principles, the resulting synthetic signals are not just visually realistic but also physiologically plausible and offer pathophysiological interpretation. This type of constrained synthesis is essential because it guarantees that the generated data, while artificial, is governed by the same biological rules as real data, thereby improving the robustness and interpretability of models trained upon it.

2.3. Foundation Models (FMs) and Large-Scale Pretraining

The latest development in PPG synthesis involves the application of large-scale Transformer-based foundation models (FMs).

2.3.1. GPT-PPG: Transformer-Based Generative Pretraining

GPT-PPG represents the first foundation-scale model in the PPG domain, adapting the high-capacity Transformer (GPT) architecture for time-series data. The model leverages self-attention mechanisms to effectively capture long-range dependencies across PPG sequences. It was pretrained on an unprecedented scale, using approximately 200 million 30-second PPG segments from multiple datasets. This massive pretraining effort enables the model to learn highly generalized PPG representations. GPT-PPG provides both generative (synthetic PPG) and predictive capabilities, achieving state-of-the-art (SOTA) performance in various downstream tasks, including AF detection and signal denoising. It can also generate patient-specific synthetic signals for data augmentation. However, the prohibitive computational and memory requirements associated with training such immense foundation models present a significant logistical barrier to entry for the majority of research groups.

2.3.2. Pulse-PPG: Emphasis on Field Generalization

Pulse-PPG addresses a critical gap identified in GPT-PPG—the reliance on highly controlled clinical data. Pulse-PPG was trained exclusively on uncataloged, raw field data collected over 100 days from 120 participants. This strategic focus on real-world variability, encompassing high noise levels and diverse contextual factors, allows Pulse-PPG to learn representations that are significantly more robust to noise than models trained solely on clinical data. This results in superior generalization and performance across both clinical and mobile health applications. This approach underscores a crucial realization: maximizing data diversity and exposure to real-world noise is essential for training universal algorithms intended for ubiquitous health monitoring. Table 1 provides a comparison of the key synthetic PPG generation methodologies discussed.

3. Applications of Synthetic PPG in Wearable Healthcare and Clinical Systems

The practical utility of synthetic PPG is most critically demonstrated in addressing the inherent data challenges faced by real-world wearable devices, where signal quality and the need for personalized data are constant issues.

3.1. Signal Quality Enhancement and Denoising

Robust performance of wearable PPG devices is contingent upon their ability to maintain signal quality despite motion artifacts (MAs), which constitute the primary source of real-world measurement error. Synthetic data generation techniques are indispensable in this domain. Generative models, such as the Recursive GAN (R-GAN) and the Deep Convolutional GAN (DC-GAN), are trained specifically using synthetic noise profiles that accurately mimic real-world MAs and environmental interference.

The R-GAN, with its recursive refinement mechanism, demonstrated impressive results, successfully reconstructing approximately 77% of corrupted PPG segments accurately. This enhanced signal quality reduced the Heart Rate Mean Absolute Error (MAE) to approximately $1.7\ \text{bpm}$ even when measured under motion. Similarly, a DC-GAN approach achieved a high correlation coefficient ($R\approx0.98$) with reference signals during high-motion interference, validating the use of generative models for real-time denoising. The efficacy of training signal reconstruction models on synthetic data arises from a unique advantage: it allows researchers to generate large, labeled datasets of corrupted signals where the "clean" ground truth is perfectly known. This is practically impossible with real-world data where the true, noise-free signal is unattainable. This capability is essential for rigorously training and benchmarking reconstruction algorithms, ensuring they perform reliably in unpredictable field conditions encountered by wearable users.

3.2. Cuffless Continuous Vital Sign Estimation

The quest for non-invasive, cuffless Blood Pressure (BP) monitoring represents a major technological driver for synthetic PPG research. BP estimation models require complex deep learning architectures to translate subtle morphological features of the PPG waveform into accurate Systolic (SBP) and Diastolic (DBP) values.

3.2.1. PPG to Arterial Blood Pressure (ABP) Translation

The PPG2ABP project employed a two-stage cascaded approach, utilizing a 1D U-Net to approximate the Arterial Blood Pressure (ABP) waveform, followed by a MultiResUNet for waveform refinement and error reduction. This model leveraged data augmentation (e.g., using the PPG-DaLiA dataset) to improve its generalizability. The resulting predictions achieved clinically relevant accuracy: MAE SBP approx 5.7 mmHg and DBP approx 3.4 mmHg, meeting British Hypertension Society (BHS) standards (Grade A for DBP).

Table 1. Comparison of Key Synthetic PPG Generation Methodologies

Methodology Class	Example Model/ Technique	Core Generation Principle	Noted Advantage	Primary Limitation
Dynamic Modeling	PPGSynth (Gaussian)	Two combined Gaussian functions + HR distribution	High flexibility/control over parameters	Lack of real-world physiological complexity
Optimization-Driven	Genetic Programming (GP)	Evolving mathematical expressions	Low computational overhead, edge-friendly	Limited waveform diversity/pathology modeling
Advanced GANs	LSM-GAN	Adversarial training + Log-Spectral Loss	Enforces spectral fidelity (frequency domain)	Computational demand, dependence on adversarial stability
Reconstruction GANs	Recursive GAN (R-GAN)	Recursive refinement of generator output	Effective artifact removal (temporal consistency)	Increased computational complexity compared to single-pass
Foundation Models	GPT-PPG	Transformer-based large-scale pretraining	Superior generalization, SOTA denoising/AF detection	Extremely high computational cost for training
Hybrid Models	Physics-integrated VAEs	Combines physiological priors with statistical modeling	Improved interpretability, physiological plausibility	Requires complex domain knowledge integration

3.2.2. Advanced BP Prediction Models

Further advancements utilize highly complex generative architectures, such as the Deep Sparse Residual U-Net (DSRUnet) combined with Squeeze and Excitation-Gated Recurrent Unit (SE-GRU) modules. This architecture is specifically optimized for capturing subtle temporal dynamics within the PPG signal, leading to state-of-the-art performance in BP prediction, achieving MAE values of $3.36 mmHg for SBP and 2.35 mmHg for DBP.

The critical requirement for non-invasive BP models is achieving stringent clinical accuracy thresholds, such as those set by BHS or the Association for the Advancement of Medical Instrumentation (AAMI). The attainment of these metrics, which validate the clinical

utility of the device, is only feasible because synthetic data allows for the massive augmentation and diversification of the training corpus. Without the capacity to supplement scarce real ICU or clinical data with high-quality synthetic signals, these complex deep learning models would be constrained.

3.3. Digital Twin (DT) Technology and Personalized Medicine

Synthetic PPG plays a pivotal role in the paradigm shift toward Digital Twin (DT) technology for predictive analytics, moving towards highly personalized medicine. By leveraging synthetic PPG generation tools, researchers can simulate the behavior of specific physiological signals (e.g., radial artery waveforms) under precisely controlled virtual conditions. This allows for the precise testing of algorithm robustness and the calibration of device parameters, such as developing low-power interpolation methods required to extract high-resolution inter-beat intervals from low-sampling-rate wearable sensors.

The integration of synthetic data within a DT framework supports sophisticated applications, including simulating disease progression and tailoring treatment plans based on an individual's virtual physiological profile. Table 2 summarizes the utility of synthetic PPG across different advanced applications.

4. Discussion, Critical Challenges, and Future Research Directions

4.1. Evaluating Physiological Realism and Standardization

Despite significant methodological progress, a primary challenge remains the assurance of physiological realism. Many synthetic signals, particularly those derived purely from mathematical models, struggle to capture patient-specific physiological nuances and the full spectrum of complex, subtle pathological patterns. Simple statistical metrics like correlation coefficients (r) or Mean Squared Error (MSE) are insufficient validation tools for determining clinical relevance or true physiological fidelity. Future research must therefore adopt more rigorous, statistically grounded evaluation.

Currently, the lack of a universally accepted framework for benchmarking and assessing the quality or clinical relevance of synthetic PPG data hinders direct, objective comparisons across different generative models. Standardization in both generation methodology and validation metrics is essential to build trust and accelerate the reliable deployment of AI systems trained on artificial signals.

4.2. Navigating the Fidelity-Utility-Privacy Trilemma

The generation of biomedical synthetic data is fundamentally constrained by a trilemma involving fidelity (how accurately it resembles real data), utility (its performance in downstream ML tasks), and privacy (protecting patient identities). Optimizing one constraint often compromises another. For instance, highly restrictive privacy protection mechanisms, such as Differential Privacy (DP), have been shown to have a detrimental effect on crucial feature correlations within the dataset. Since physiological signals rely heavily on complex, nonlinear relationships.

4.3. Key Future Research Directions

1). Physics-Informed Generative Models: Further research is needed to embed more complex physiological constraints into deep learning architectures (GP+CNN architectures).

2). Open-Access Benchmarking: The establishment of large-scale, open-access synthetic PPG datasets is necessary to facilitate standardized research and allow for robust, equitable benchmarking of new generation and application models across the entire biomedical community.

3). Comprehensive Clinical Validation: Rigorous, large-scale validation of synthetic

data is required in practical, real-world healthcare applications, focusing on measures of effectiveness, safety, and clinical outcomes in areas such as remote monitoring and wearable diagnostics.

4). Ethical and Regulatory Frameworks: Future work must address the evolving ethical, legal, and regulatory aspects of using synthetic data in biomedical research and practice, ensuring public confidence and patient safety in AI systems trained on non-real signals.

Table 2. Utility of Synthetic PPG in Downstream Applications

Application	Task	Model/Technique Used	Synthetic Data Role	Performance Metric Achieved (MAE/ Accuracy)	Clinical Translation
Signal Quality Enhancement	Motion Artifact Removal	R-GAN, DC-GAN	Training Denoising models on known ground truth	HR MAE approx1.7 bpm	Ensures real-time accuracy of wearable devices
Cuffless BP Estimation	SBP/DBP Prediction	DSRUnet + SE-GRU	Massive data augmentation/ diversification	MAE SBP approx 3.36 mmHg DBP approx 2.35 mmHg	Enables clinical grade cuffless continuous monitoring
Cuffless BP Estimation	ABP Waveform Translation	1D U-Net (MultiRes UNet) on (PPG-DaLiA)	Data augmentation for generalizability	MAE DBP 3.4 mmHg (BHS Grade A)	Key step towards cuffless continuous monitoring
AF Detection	Alleviating class imbalance of AF signals	LSM-GAN Augmentation	Simulating rare pathological events	Improved AF detection accuracy by 12%	Crucial for developing screening tools for low-prevalence pathologies

5. Conclusion

This academic review has thoroughly detailed the motivation, generative methodologies, and critical applications of synthetic Photoplethysmogram (PPG) data. The transition from rudimentary Gaussian dynamic models to sophisticated Generative Adversarial Networks (GANs) and large-scale Foundation Models like GPT-PPG has successfully addressed the primary constraints of data scarcity, high annotation costs, and patient privacy

[1, 3]. Synthetic data has proven to be functionally essential, particularly in wearable healthcare, by enabling the training of robust signal reconstruction algorithms (R-GAN), facilitating clinical-grade accuracy in cuffless Blood Pressure estimation (DSRUnet), and providing vital data augmentation to combat class imbalance in pathology detection (LSM-GAN).

Despite these transformative strides, the field must confront persistent challenges relating to the gap between high correlation metrics and true physiological fidelity, the absence of standardized validation frameworks, and the complex engineering required to resolve the trade-off between privacy mechanisms and data utility [1, 8]. The future trajectory of non-invasive vital sign identification systems rests upon the joint progression of clever noise reduction techniques and the prudent use of high-quality synthetic data. By prioritizing the development of physics-informed, generalizable, and privacy-aware generative models, synthetic PPG will be cemented as an indispensable tool for accelerating biomedical research and the clinical introduction of next-generation health monitoring solutions.

References

1. Kim K.B., Baek H.J., Photoplethysmography in Wearable Devices: A Comprehensive Review of Technological Advances, Current Challenges, and Future Directions, Electronics. Accessed: Oct. 23, 2025. [Online]. Available: https://www.mdpi.com/2079-9292/12/13/2923
2. Vos G. et al., A Statistical Approach for Synthetic EEG Data Generation, arXiv. Accessed: Oct. 23, 2025. [Online]. Available: https://arxiv.org/html/2504.16143v2
3. Adams T. et al., On the Fidelity versus Privacy and Utility Trade-Off of Synthetic Patient Data, medRxiv. Accessed: Oct. 23, 2025. [Online]. Available: https://www.medrxiv.org/content/10.1101/2024.12.06.24317239v2
4. Cordeiro M.C et al., A synthetic data-driven machine learning approach for athlete performance attenuation prediction, PMC. Accessed: Oct. 23, 2025. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC12148907/
5. Kataria S., Wu Y., Hu X., Foundation Models for Physiological Data, SIAM.org. Accessed: Oct. 23, 2025. [Online]. Available: https://www.siam.org/publications/siam-news/articles/foundation-models-for-ph ysiological-data/
6. Song R., Pulse-PPG: An Open-Source Field-Trained PPG Foundation Model for photoplethysmography. Accessed: Oct. 23, 2025. [Online]. Available: https://renchengsong.github.io/papers/PulseGAN.pdf
7. Saratkar S.Y. et al., Digital twin for personalized medicine development, Frontiers. Accessed: Oct. 23, 2025. [Online]. Available: https://www.frontiersin.org/journals/digital-health/articles/10.3389/fdgth.2025.158 3466/full
8. Adams T. et al, On the fidelity versus privacy and utility trade-off of synthetic patient data, PMC. Accessed: Oct. 23, 2025. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC12059695/

Machine Learning in Healthcare Staff Training and Assessment

R Rishab, Sivakumar Rajagopal

Vellore Institute of Technology, Vellore, Tamil Nadu, India
rishab.r2022@vitstudent.ac.in, rsivakumar@vit.ac.in

Abstract. Conventional approaches to training in healthcare are limited by resources and self-reporting. As a result, there is a sizable gap in measuring the effectiveness of training in context to affective employee behaviors and organizational outcomes long term. Machine learning (ML) is a data-driven solution to this issue that allows for a more scalable approach to training assessment. ML models can objectively analyze behavioral data, getting away from surveys, to potentially predict training success. Using ML could enhance how training is evaluated later in service quality. Future studies are necessary to explore the viability and validation of ML methodologies to steer the educational paradigm away from conventional approaches.

Keywords. Healthcare Staff Training, Machine Learning (ML), Training Assessment, Training Effectiveness, Training Modernization.

1. Introduction

In recent years, the way organizations train and develop staff and employ learning practices has changed dramatically from relatively fixed and standardized models and programs to more individualized, flexible, and data-focused approaches. Organizations are increasingly under pressure to prepare their employees to develop and incorporate new skills continuously and engage in learning that is adaptable to change, in a time of rapid technological change, globalization, and increasingly disruptive change to organizations and entire sectors, such as the business and healthcare sectors. Traditional training programs and practices, despite their value and place, do not adequately provide individuals and organizations with adequate or proper training experiences to facilitate continuous and effective skill development and learning, cater to a variety of employee learning needs, meet the needs of today's work complexities, and demonstrate measurable results.

The emergence of machine learning (ML) and artificial intelligence (AI) opens up potential new and incredible ways that organizations can think about the design, delivery, and evaluation of training experiences. Both of these technologies can enable the creation of intelligent learning systems that can analyze learner behaviors, recommend personalized content, provide predictions regarding skill gaps, provide real-time feedback—even the most basic of these uses alone surpass conventional evaluation methodologies, such as surveys or performance appraisals. Even with all of this great potential, ML and AI applications, especially in Human Resource Management (HRM), are in their infancy.

2. The Reality of Training: Insights from Non-ML Studies

A considerable amount of prior research on training and evaluation of staff, and for effects of training on staff, continues to use non-machine learning (ML) approaches. These include as rule-based systems, statistics, and heuristics-based approaches, which have had

significant achievements in understanding training effectiveness and developing valid performance standards. Unfortunately, traditional approaches typically lack the ability to address large-scale, heterogeneous, or evolving datasets, ultimately obscuring critical limitations observing distinction in learning effects absent more nuanced effects or longitudinal impacts, or adjust to complex real-world tasks and domains. Machine learning-based approaches can address these limitations. Machine learning typically involves ML learning from data, adapting to changes in data patterns, and continue to improve its ability to predict better than traditional approaches. Machine learning has the potential to address traditional approaches limitations effectively:

- Methodological Flaws: A large proportion of studies use assessment tools such as survey, semi-structured interview, and self-reported questionnaire-type evaluations to assess training programs. While assessment tools are convenient to administer and provide immediate data, they also have substantial limitations. First, they rely heavily on subjective perceptions that may not adequately reflect any changes to skills or behavior in the workplace. Second, most studies do not have control groups or some other standardized comparison, which decreases the draw of their findings. Finally, few studies look at the evaluation beyond the training immediately or shortly after to examine the long-term impact on behavior, organizational performance, or patient outcomes. This is particularly problematic in healthcare and high-stakes professional settings, where improvement in behavior over time is the eventual indicator for successful training.

- Practical Gaps in Training: Studies have also established ongoing gaps between training and workplace needs. For example, survey data from certified professionals indicate that while the quantity of training opportunities has increased in recent years, many organizations are still using simple instructional strategies including verbal instruction, observation, and modeling. These strategies can be useful for some contexts, but they are inadequate for developing the advanced competencies necessary in complex professional environments. For example, just under half of supervisors indicated that they had undertaken formal training in supervisory or leadership development skills, suggesting a systemic gap in preparation for management. Interventional studies also provide evidence of this training gap. One interventional study on trauma-informed care training, for example, found that even a brief four-hour training resulted in significant improvements in participants' self-reported competence. Moreover, this highlights a large reliance on subjective outcome measures, which implies an increased need for objective assessment measures that provide us with more rich, rigorous, valid, and comprehensive outcome measures across supervision and training.

The existing literature indicates that many healthcare training programs continue to utilize more conventional approaches, such as verbal instruction, observation, and modeling. Although they provided a solid foundation, these methods, and their evaluation, are not scalable or support adaptability in a time of increasing complexity. While intervention studies on trauma-informed care, for example, report short-term increases in self-perceived competence, they also tend to diminish the credibility of their findings through sample size or sole reliance on subjective, self-reported data. More importantly, these findings continue to leave unanswered questions about lasting, objective changes in practice patterns as a result of the focus on perceived competence post-intervention. These examples reinforce a common critique found in the healthcare training literature: while evaluations using traditional evaluation frameworks may indicate short-term improvements, they fail to assess long-term changes in behavior or improvement in objective, standardized results at the level of patient care practices. As such, this highlights the urgent need to utilize more scalable, data- and objective-based evaluation frameworks in practice, such as those based on machine-learning

(ML) methodologies, that will deepen effectiveness, measure longitudinal outcomes, and be adaptable to complex and dynamic healthcare practice.

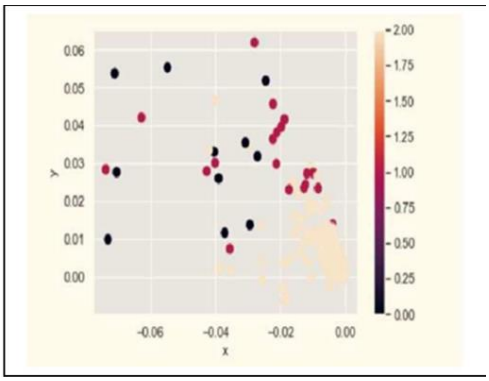


Figure 1: Singular Value Decomposition (SVD) algorithm. personalize competency management by providing tailored training suggestions to employees

3. Machine Learning for Training Personalization and Recommendation Systems

Current research indicates a better movement towards tailoring training using machine learning (ML), departing from the model of a standardized, one-size-fits-all approach. An example to note are hybrid recommender systems that blend content-based filtering with collaborative filtering. In these types of recommender systems, content-based filtering takes an employee's profile (career goals, interests) and applies those attributes to a training resource; whereas, collaborative filtering uses employee profiles to recommend resources based on the behaviors and preferences of employees who shared similar profiles, and may utilize algorithms such as Singular Value Decomposition (SVD). This all-encompassing method has a greater benefit than simply providing a list of courses. When using an optimized ranking algorithm, these systems can reduce the time it takes to train a model by ~50% while increasing accuracy in the recommendations. Furthermore, they provide management with strategic insights, such as identifying trending skills and using that information to allocate budgets for training investment. Taken together, with a technology-enabled capability, these systems can help the organization build training frameworks that are efficient and aligned to both employee development and the needs of the organization.

The productivity improvements of these systems go far beyond basic course matching. Research shows that when systems use a strong ranking algorithm in their overall ML frameworks, these systems can reduce the time it takes to train the model and refine recommendations by around 50%, while improving personalization. In addition to helping employees optimize their own learning, these systems can also yield valuable management insights, such as which skills are growing more important in the workforce, identifying training demands that could potentially be addressed by a formal learning resource, and informing how budgets should be allocated to workforce development. In this way, ML-mediated personalization helps achieve both employee and organizational objectives. The implications within healthcare are particularly profound. A hybrid recommendation system based on ML could tailor learning and professional development opportunities for clinical personnel according to people's specialty, years of experience, and career trajectory. For example,

surgeons might receive a recommendation for an advanced module of learning about new medical devices, while nurses working in specialized wards would be recommended learning about how to communicate with patients more effectively. The prospect of achieving a 50% reduction in learning and training time is invaluable in a hospital context, where the clinical environment is high-risk with few personnel to spare, and the demand for upskilling, quickly, is great.



Figure 2: Implementation of Machine learning in Human resource management (HRM)

4. Machine Learning for Performance Measurement and Outcome Prediction in Healthcare

Machine learning (ML) may provide a more neutral and holistic lens for evaluating training in healthcare than traditional evaluation methods. In traditional approaches, such as administering surveys or supervision evaluations, healthcare professionals have limited and often subjective perspectives on training's effectiveness. However, ML has the potential to analyze large and heterogeneous data, exposing complex non-linear patterns in the relationships between training interventions, employee behavior, and patient outcomes. This section reviews three important areas of ML value to date in healthcare: behavioral outcomes, organizational outcomes, and drivers of engagement.

- Behavioral Outcomes: The behavior and attitudes of employees are critical factors in delivering healthcare services, because of impacts on such factors as staff retention, team dynamics, and ultimately, patient care. Therefore, methods such as Machine Learning (ML), including Random Forest Regression, have been utilized to predict how training programs, or interventions, will influence key behavioral metrics such as job satisfaction, or organizational citizenship behavior. Evidence of these predictive validity concerns, in similar studies outside of healthcare environments, is supportive (for instance, R^2 values exceeding .85), indicating ML methodologies for enhancing healthcare workforce capabilities will yield benefits.

In a hospital context, models can be trained incorporating records of participants in training programs, along with evaluations of performance, and results of employee surveys. Predictive outputs could report the likelihood of improved job satisfaction as an outcome for an employ an employee of a specific training program, which would help administrators identify which training programs for skills, competencies, or capabilities deliver the most long-term impact on workforce stability and morale. Additionally, the models often pick up subtleties and relationships that exist between variables (interacting variables like workload,

or leadership style) that traditional statistical technique may not detect, especially in studies based on surveys or ratings (although no compromise in predictive validity; among other attribution variables).

- Outcomes in Organizations: In addition to addressing the individual employee level, ML will provide a mechanism for linking training investments to organizational outcomes. Linking training investments will become vital as an Artificial Neural Network (ANN) has been applied to an organizational context to model complex organizational systems where multiple "inputs" related to human and technology performance combine to effect outcomes such as productivity, service quality, or efficiency. In the health context, ANNs could be specified to evaluate how combinations of factors such as staff training, use of technology, and resource allocation/evaluation, could systematically influence organization-wide outcomes such as patient throughput, wait times, and/or readmission rates. For example, through predictive modeling, a hospital could identify if the level of staff training within an organization to the introduction of new EHR would yield measurable improvements on operational efficiency or reduction in error rates. Evidence suggests that technology alone, when trained and utilized with Human Resource Development, would account for over 70% of performance gains. This implies that the alignment of Human Resource Development, by training and organizational technology, could be highly valued and beneficial on improving health care outcomes. For those decision-makers that would seek to evaluate presenteeism for healthcare services, the predictive results will represent meaningful value. Through prediction modeling, investment firms could evaluate Virginia returned value on investment (e.g., ROI) in link to some formalized professional development, and to increase efficiency as well as to solidify program professional development. In addition, travel costs would fortify the training system's rationale and personnel development in support of the hospitals' strategy to enhance performance and become more competitive offered in the context of advantage of technology.
- Factors of Engagement: Machine Learning has also been used to determine the vital factors of staff engagement and when training plays a role, induces the association, i.e., an engagement factor will mediate the relationship between training and outcomes, including patient satisfaction and quality of care. In health care, engagement is related to decreasing burnout, improving team work, and increasing safety practices. ML studies utilizing tree-based models, such as Gradient Boosted Trees (GBT) & Random Forest, have consistently identified safety culture and teamwork as the most potent predictors of staff engagement. In fact, the evidence provides a clear and data grounded evidence case as to why training programs evaluate the organizational culture (ensuring a strong culture of safety) and interprofessional collaboration (teamwork). At the least, if engagement factors can be captured, this too will have the most predictive effect. Because there are high-impact engagement factors, organizations should focus on them to the maximize the effect of their training on staff wellbeing as well on patient reported outcomes. Tree based model outputs are also a benefit because of feature ranking, which allows the results to be interpretable for non-technical stakeholders. Leaders of hospitals will be able to not only see what the most important factors are, but also what factors are important in each context and their relative ranking.

5. Reframing Workforce Training in Healthcare: Machine Learning as a Strategic Asset

Research indicates a significant shift in the conceptualization, implementation, and assessment of staff training in healthcare organizations, driven primarily by the rise of machine learning (ML). This shift is not solely technological but represents a shift toward, rather than

being reactive and standardized, proactive, adaptive, and evidenced-based workforce development. This review identified several significant trends.

- From Subjective to Objective Measurement

Healthcare training evaluation has traditionally relied on subjective approaches such as self-reporting surveys, interviews, and post-training assessments. However, ML models offer transition to objective and data-driven assessment. In particular, ML systems have the potential to incorporate heterogeneous data sources such as electronic health records, competencies, and patient outcomes, to assess the impact of training interventions in standings that can be required as rigorous. This allows for the measurement of fine-grain change of competencies (e.g., communication, safety compliance, procedures accuracies) as markers to link to organizational outcomes in relation to training interventions likely to improve error rates and care quality and productivity. Evidence-based assessment provides healthcare systems with a professor-reliable method for connecting adult workforce training directly to outcomes that represent patients and the delivery system.

- From One-Size-Fits-All to Personalized Learning

A key shift in the literature is the transition from standardized, "one size fits all" training models to personalized ML-enabled learning paths. By analyzing staff characteristics such as personal demographics, previous learning experience, clinical specialty, and career goals, ML algorithms can recommend personalized training. In practice, this could mean pointing a novice nurse to modules on patient safety and medication management while pointing a more senior clinician to advanced leadership or subspecialty training. Personalized training increases the relevance and effectiveness of instructional materials, but it can also increase staff engagement, retention, and professional development—all issues important to addressing healthcare workforce shortages and burnout.

- From Automation to Interpretability

The initial uses of AI in healthcare workforce management were mainly administrative, focusing on resume screening or compliance monitoring. However, recent studies highlight the need to focus on interpretability and transparency of ML-driven training recommendations. Models like BERT and LSA have been valued as both predictive models, but also for their explainability. This represents a considerable step forward in the ability of healthcare administrators and educators to see the basis for training assignments or why performance may have flagged; therefore, improving trust, accountability, and adherence to ethical and regulatory considerations. Interpretability is the first step in the movement from automation to supportive decision making, allowing ML to be integrated into healthcare training in both responsible and useful ways.

- Technology as a Strategic Asset

In conclusion, the studies we reviewed bring us to common ground in recognizing ML as a strategic asset to develop the healthcare workforce. Rather than seeing ML-enabled systems solely as efficiency tools, those in healthcare contexts increasingly see ML-enabled systems as contributing to organizational performance. In aligning front-line staff training programs to measurable improvements (e.g., patient safety scores, improved wait times, and enhanced competencies), ML recalibrates training as an investment to support broader institutional measures. This strategic alignment frames workforce development as a central characteristic of healthcare quality improvement, accreditation, and patient-centered care; workforce development is not viewed as a secondary or administrative function in healthcare contexts.

6. Future Scope

Applying machine learning (ML) to the training of healthcare personnel has substantial potential, but future work needs to focus on closing the gap between the promise of conception and actual practice. Many existing studies emphasize the potential of ML to improve the effectiveness of training, but the evidence is mostly limited to theoretical models or pilot projects. To demonstrate the effectiveness of ML in 'real-world' training interventions, there needs to be large-scale effectiveness through randomized controlled trials and longitudinal studies, not only to show the effectiveness of ML on the improvements of knowledge and skills, but also for the downstream effects of clinical decision-making, patient safety, and organizational efficiency. Demonstrating those outcomes at scale will be an important step in facilitating uptake and institutional credibility. Just as vital is the need to tackle ethical, transparency, and fairness issues when ML is applied to staff development. Black-box algorithms carry risks of unfairness and unequal treatment, which may disadvantage staff on grounds of gender, ethnicity, or professional experience. To minimize such risks, the design of explainable, fairness-conscious ML systems with explicit reasoning behind recommendations should be a priority in future research. Creating governance structures, bias-tracking procedures, and ethical design guidelines will ensure that ML-facilitated training fosters equality and sustains trust among healthcare providers and administrators.

Lastly, ongoing work needs to concentrate on developing integral and sustainable learning environments that blend ML innovation with tried and tested concepts of healthcare education. Adaptive platforms with ML, microlearning tools, and live performance feedback have the capacity to become critical enablers in solidifying both entry-point and ongoing learning. These technology solutions need to be integrated within competency-based systems and adult learning models to guarantee systematic growth through career phases. In addition, researchers must explore the organizational-level impacts of ML-facilitated training, such as its role in fostering interdisciplinary teamwork, readiness for accreditation, and overall quality of care. By linking personalized learning with institutional objectives, ML may transform from a mere technological aid to a strategic force for workforce development and patient care.

7. Conclusion

The digital transformation of training for healthcare staff is no longer a remote possibility-it is an unfolding reality. Machine learning (ML) arose as an instrumental enabler shifting the traditional approaches toward data-driven, adaptive, and outcome-based strategies. ML has, in fact, addressed long-standing limitations in workforce development by enabling changes in phenomena from subjective assessments toward objective measurement; in effect, standardized programs are toward an individualized pathway; and administrative automation is toward interpretability. These innovations influence workforce development directly in such a way that staff training becomes a driver for patient safety, quality of care, and efficiency of healthcare to the extent that it is no longer a mere period of administrative activities.At the same time, all the challenges concerning the adoption, validation, and ethical issues have to be taken into consideration in the ML integration into the training in healthcare. Initial studies have demonstrated strong potential, while extensive empirical studies still need to be pursued to establish the real-world efficacy of the ML-based approaches. The issues of algorithmic bias, unfairness in transparency, and prevalent dilemma of fairness in providing opportunities in training programs form a major concern; hence, the notion of justice and accountability must be incorporated into the system design stage. If these issues are left unresolved, the mistrust or ill-sided outcomes would prevent the potential of ML from being realized in an institutionalized setting of healthcare.

Future looking experiences in successful workforce development for healthcare may be dependent on the alignment of technological innovations with established educational theories and institutional considerations. The embedding of adaptive learning systems into competency-based systems, fostering a form of lifelong professional development, and associating outcomes with organizational goals thereby lead to a self-sustaining setup for workforce training. Different from improving ML systems, in the end, ML can really make training faster and efficient.

References

1. S. Sinclair, J. Kondejewski, P. Jaggi, L. Dennett, A. L. Roze des Ordons, and T. F. Hack, "What Is the State of Compassion Education? A Systematic Review of Compassion Training in Health Care," *Academic Medicine*, vol. 96, no. 7, pp. 1057–1070, Jun. 2021, doi: https://doi.org/10.1097/acm.0000000000004114.
2. T. W. Foong, E. S. Hui Ng, C. Y. Wee Khoo, B. Ashokka, D. Khoo, and R. Agrawal, "Rapid training of healthcare staff for protected cardiopulmonary resuscitation in the COVID-19 pandemic," *BJA: British Journal of Anaesthesia*, May 2020, doi: https://doi.org/10.1016/j.bja.2020.04.081.
3. T. Barnes and M. Wilkie, "A learning process to deliver virtual staff training involving patients in shared haemodialysis care," *Clinical Kidney Journal*, vol. 16, no. 1, pp. 48–56, 2023, doi: https://doi.org/10.1093/ckj/sfad067.
4. K. M. S. Alharbi and S. A. S. Aloyuni, "The importance of training and development of employees in improving the quality of health services," *International journal of health sciences*, vol. 7, no. S1, pp. 2190–2201, Aug. 2023, doi: https://doi.org/10.53730/ijhs.v7ns1.14473.
5. P. McGee-Vincent *et al.*, "Training Staff Across the Veterans Affairs Health Care System to Use Mobile Mental Health Apps: A National Quality Improvement Project," *JMIR Mental Health*, vol. 10, p. e41773, Jan. 2023, doi: https://doi.org/10.2196/41773.
6. R. F. Rad *et al.*, "Comparative study of virtual and face-to-face training methods on the quality of healthcare services provided by Kermanshah pre-hospital emergency staff (EMS): randomized educational Intervention trial," *BMC Medical Education*, vol. 22, no. 1, Mar. 2022, doi: https://doi.org/10.1186/s12909-022-03277-y.
7. N. Vera San Juan *et al.*, "Training and redeployment of healthcare workers to intensive care units (ICUs) during the COVID-19 pandemic: a systematic review," *BMJ Open*, vol. 12, no. 1, p. e050038, Jan. 2022, doi: https://doi.org/10.1136/bmjopen-2021-050038.
8. C. A. Surr *et al.*, "Components of impactful dementia training for general hospital staff: a collective case study," *Aging & Mental Health*, vol. 24, no. 3, pp. 511–521, Dec. 2018, doi: https://doi.org/10.1080/13607863.2018.1531382.
9. M. Lamberson *et al.*, "Assessing Emergency Department Staff Knowledge, Competency, and Implementation of Pre- and Post-Trauma-Informed Care Training," *Journal of Emergency Nursing*, vol. 51, no. 1, pp. 145–157, Jan. 2025, doi: https://doi.org/10.1016/j.jen.2024.09.010.
10. R. Al-Nammari, M. C. E. Simsekler, A. F. Gabor, and A. Qazi, "Exploring Drivers of Staff Engagement in Healthcare Organizations Using Tree-Based Machine Learning Algorithms," *IEEE Transactions on Engineering Management*, vol. 70, no. 8, pp. 2988–2997, Aug. 2023, doi: https://doi.org/10.1109/tem.2022.3209879.
11. H. Aguinis and J. Burgi-Tian, "Talent management challenges during COVID-19 and beyond: Performance management to the rescue," *BRQ Business Research Quarterly*, vol. 24, no. 3, pp. 233–240, Apr. 2021, doi: https://doi.org/10.1177/23409444211009528.
12. A. Blackman, F. Digennaro Reed, T. Erath, and A. Henley, "A Survey of Staff Training and Performance Management Practices: An Update," vol. 16, pp. 731–744, 2023, doi: https://doi.org/10.1007/s40617-022-00762-0.
13. S. Basnet, "Artificial Intelligence and Machine Learning in Human Resource Management: Prospect and Future Trends," *International Journal of Research Publication and Reviews*, vol. 5, no. 1, pp. 281–287, Jan. 2024, doi: https://doi.org/10.55248/gengpi.5.0124.0107.

14. S. Sarangarajan, C. Kavi Chitra, and S. Shivakumar, "Automation of Competency & Training Management using Machine Learning Models," *IEEE Xplore*, Feb. 01, 2021. https://ieeexplore.ieee.org/abstract/document/9513995 (accessed Dec. 15, 2022).
15. Y. Sun and H. Jung, "Machine Learning (ML) Modeling, IoT, and Optimizing Organizational Operations through Integrated Strategies: The Role of Technology and Human Resource Management," *Sustainability*, vol. 16, no. 16, pp. 6751–6751, Aug. 2024, Available: https://www.mdpi.com/2071-1050/16/16/6751
16. A. Gupta, A. Chadha, V. Tiwari, A. Varma, and V. Pereira, "Sustainable training practices: predicting job satisfaction and employee behavior using machine learning techniques," *Asian Business & Management*, Jun. 2023, doi: https://doi.org/10.1057/s41291-023-00234-5.
17. S. Garg, S. Sinha, A. K. Kar, and M. Mani, "A Review of Machine Learning Applications in Human Resource Management," *International Journal of Productivity and Performance Management*, vol. 71, no. 5, pp. 1590–1610, Feb. 2021, doi: https://doi.org/10.1108/ijppm-08-2020-0427.

Recent Trends in Organizational Structures: Towards Agile and Digitally-Enabled Enterprises

Ayushmaan Panda, Sivakumar R

Vellore Institute of Technology, Vellore, Tamil Nadu, India
ayushmaan.panda2022@vitstudent.ac.in, rsivakumar@vit.ac.in

Abstract. Organizational structures are undergoing profound shifts driven by digital transformation, Industry 4.0 technologies, and the new dynamics of remote and hybrid work. Flattening hierarchies, less formalization, and the rise of cross-functional and self-managing teams along with platform-driven collaboration are all trends highlighted by recent empirical studies. More agility, customer centricity, and collaborative efficiency are being made possible by digital technologies, such as artificial intelligence (AI), the Internet of Things (IoT), and sophisticated communication platforms. These technologies are also forming new patterns of interaction and decreasing the need for strict managerial layers. Simultaneously, silo effects during remote work, cultural resistance to change, and employee fatigue risks present challenges in organizations that are constantly changing. To solve these problems, organisations must integrate digital tools, adaptive leadership, and robust governance frameworks in a balanced way to maintain employee well-being and innovation.

Keywords. Organizational structures, digital transformation, Industry 4.0, agility, self-management, collaboration networks.

1. Introduction

Effective departmental communication is essential for timely decision-making, coordination, and performance in large, complex organisations, especially in healthcare systems. Communication silos, in which teams or units function independently with little interdepartmental interaction, are a persistent problem for many of these systems. Over time, these "dynamic silos" change as a result of shifts in the workload, organisational structure, digital infrastructure, or crises like the COVID-19 pandemic.

Collaboration can be severely hindered by silos, which can result in duplicate work, slower reaction times, and lower operational efficiency. Because traditional organisational restructuring techniques are expensive and time-consuming, simulation-based methods are a desirable substitute for testing communication improvement strategies prior to actual implementation. Using actual communication data, this study suggests a simulation framework for examining and dismantling dynamic silos.

The framework uses social network analysis (SNA) metrics to quantify silos, models intra- and inter-team communication as a network graph, and simulates three different organisational intervention types to assess how well they enhance cross-team collaboration. For organisational leaders who want to improve coordination without upending current structures, particularly in hospitals, the findings offer practical insights.

2. Literature review

Network theory, empirical measurement, and intervention design must all be integrated to comprehend how organisational communication structures develop, endure, and change. The literature reviewed here includes case studies of organisational design changes, theoretical and methodological work on network modification, simulation techniques for testing controls and interventions, and empirical studies of communication networks. When combined, these studies provide a conceptual framework for modelling and assessing interventions aimed at dismantling dynamic silos.

2.1. Empirical evidence of dynamic silos and organizational communication

The main empirical finding that drives this work is given by Yang et al. (2021): Intra-organizational communication structures became more modular during the COVID-19 pandemic, displaying what the authors refer to as "dynamic silos" (3). As teams consolidated around urgent operational needs, their longitudinal analysis of organisational message traffic revealed a decrease in cross-group connectivity and an increase in within-group ties. This temporal increase in modularity encourages measurement (e.g., modularity, cross-team edge ratios) as a means of identifying new silos because it links crisis situations and remote work arrangements to structural communication fragmentation.

This empirical tradition is extended into healthcare contexts by Patterson et al. (2013), who demonstrate how network topology maps to operational pressures and coordination patterns through a network analysis of team communication in a busy emergency department (7). Their research supports the usefulness of SNA in identifying coordination bottlenecks in hospitals and shows how high-stakes settings generate unique network signatures (such as concentrated broking roles and tight intra-team clustering). Yang et al. and Patterson et al. jointly defend the use of network metrics as operational indicators of silo strength as well as the domain relevance (healthcare).

2.2. Reviews and typologies for network alteration

By creating a typology of network alteration strategies, Bunger et al. (2023) offer a crucial conceptual framework (2). Their scoping review categorises intervention mechanisms and empirical contexts and summarises how interventions work (change environment, change composition, create ties, strengthen existing ties, change node behaviour). Because it explains the levers that organisations can pull (structural vs. behavioural) and links them to anticipated effects on diffusion, broking, and modularity, this typology is helpful when creating simulation interventions. It also draws attention to a significant methodological flaw: although many strategies are conceptually proposed, it is uncommon to conduct comparative empirical tests of different strategies on the same baseline network.

By formalising network modification operations, such as adding or removing edges, changing weights, or changing node attributes, and examining their theoretical effects on network function, Everton, Carley, and Reddy (2016) provide a contrasting methodological viewpoint (4). This work supports the computational interventions used in this study by providing operational primitives (such as adding bridging edges and increasing tie probabilities) that are directly translatable into simulation code, despite being presented in a workshop setting.

2.3. Case studies and organizational design dynamics

A recent case study of intentional "silo busting" initiatives in an organisational setting is presented by De Waal & van den Berg (2025) (1). Because it documents actual managerial practices (liaison roles, structured cross-team rituals, and governance changes) and connects them to quantifiable collaboration outcomes, their work is especially instructive for the current

project. The case study grounds simulation scenarios in realistic constraints by offering empirical detail regarding implementation challenges and trade-offs (e.g., limited scale of role appointments, varying participation rates).

Nugroho & Fitriani (2023) and Oldham & Daubner-Siva (2023) study organisational evolution in digital startups and structural shifts towards self-managing forms, respectively (8,5). Both stress how organisational culture and operational procedures are impacted by structural reconfigurations. For example, introducing cross-functional squads or flattening hierarchies can increase agility, but results are contingent on conformity to organisational norms. These studies warn that if cultural fit is neglected, interventions that work in short-term experiments may have unexpected consequences. Simulation studies may raise this concern, but they cannot fully address it without qualitative validation.

2.4. Agility, diffusion, and technological change

The connection between digital transformation and organisational agility is examined by Wu et al. (2023) (9). Their empirical results, albeit from the public sector, show that agility mediates the successful adoption of digital tools, which in turn influences the evolution of communication patterns. This study emphasises the need to take technological affordances into account when considering interventions. For instance, the addition of liaison roles works better when digital platforms allow for continuous cross-team channels.

In their discussion of how organisational form and hierarchy affect the diffusion of practices and structures, Saka-Helmhout et al. (2022) situate organisational diffusion processes within a larger theoretical framework (6). Their contribution implies that the speed at which interventions spread throughout an organisation is conditioned by structural constraints (authority, formal reporting lines), which is another justification for measuring both local (node-level) and global (modularity, path lengths) network properties.

2.5. Simulation and real-time modeling approaches

The use of real-time simulation environments to assess controls in intricate technical systems is illustrated by Ogilvie et al. (2020) (10). The methodological lesson is applicable even though their focus is shipboard power systems: safe, repeatable experiments to test interventions prior to actual deployment are made possible by simulation environments that faithfully replicate system topology. Similarly, the method of using an empirical baseline network (such as Enron or hospital proximity data) and applying controlled perturbations to compare outcomes across strategies is informed by network simulation studies and Zhao et al.-style multi-relational simulations (quoted in previous conversations).

2.6. Synthesis: common findings and methodological gaps

Several recurring themes show up in these works. First, temporal analyses show that crises and the adoption of digital tools speed up changes in modularity, and modularity and community detection are trustworthy markers of siloing (3,7). Second, although culture and sustainability concerns limit their long-term success, structural interventions (such as adding bridging roles or ties) frequently result in larger short-term reductions in modularity than diffuse behavioural interventions (1,2,4). Third, while modelling and simulation offer a low-cost way to compare interventions, there aren't many standardised comparative experiments in the literature that systematically test different strategies using the same empirical baseline (2).

There are still significant gaps. External validity and domain-specific calibration are limited by the lack of publicly accessible, high-resolution hospital communication datasets (sender/receiver/timestamp/team) appropriate for intervention testing. Furthermore, there is a gap between structural change and actual performance because most studies use modularity and related SNA metrics to measure impact while undervaluing operational outcomes (such as decision latency, task completion, and patient outcomes). Last but not least, not many studies

completely incorporate temporal multilayer networks or take into consideration tie decay/formation dynamics in their simulations, despite the fact that these dynamics are essential for comprehending how long an intervention's effects last.

2.7. Positioning the present study

Several of these gaps are directly addressed in the current simulation study. In addition to comparing interventions quantitatively using modularity and inter-team ratios, it presents repeatable simulation code that operationalises theoretical network modification primitives (4) and applies a set of empirically motivated interventions (liaisons, rituals, and rotations) on an actual communication baseline (Enron) modified to a hospital-like topology. The study contributes to the literature by providing a comparative, repeatable, and practically grounded experiment that aids managers in prioritising interventions with predictable short-term effects. It also lays the groundwork for future research that can integrate actual hospital logs and outcome-oriented evaluation, despite acknowledging the limitations of synthetic team assignment.

3. Methodology

3.1. Network Construction

The NetworkX library was used to create a weighted undirected graph where Employees or units are represented by nodes. Email correspondence between two people is represented by edges. Communication frequency is indicated by edge weights. To differentiate between intra-team (communications within the same department) and inter-team (communications across departments), sender and recipient team IDs were saved for each edge.

3.2. Silo Detection using Community Analysis

The network's natural clusters were found using the Louvain community detection algorithm.

The degree of siloing was measured by the modularity (Q) score: $Q > 0.5$, Strong silos (weak cross-links, dense internal links) Q>0.5. Q<0.3 means a communication network that is well-integrated. Additionally, network-level metrics were calculated, including density, average degree, and inter-/intra-team edge ratios.

3.3. Simulation of Communication Interventions

The main experimental element of this investigation is the simulation phase.
After building the communication network and determining the baseline modularity, a number of network modification techniques were used to simulate actual managerial actions meant to break down communication silos. These interventions were created to address the behavioural as well as structural components of improving communication in complex organisations, like big businesses or hospitals.

Rationale and Objective

Workload, shifts in leadership, project priorities, and crisis situations all influence how people communicate in dynamic organisations. Dynamic silos, in which departments grow more isolated over time, are frequently the result of this evolution.

Thus, this simulation framework has two goals:

1). In order to simulate interventions as network changes that could be applied in a practical situation;

2). To use metrics for modularity and interconnectivity to quantitatively assess their effect on network cohesion.

The guiding hypothesis is that modularity (Q), a measure of improved collaboration, will decrease when more cross-team links are added or the likelihood of interdepartmental communication increases.

3.4. Evaluation Metrics

Network Density (D): overall connectivity of the organization.

Average Degree ($\bar{k}$): communication frequency per node.

Intra- vs Inter-Team Edges (E_{in} / E_{out}): ratio indicating silo intensity.

Modularity Change (ΔQ): measure of improvement; reduction in modularity = reduced silo strength.

4. Results

The extended Enron email dataset, which was designed to mimic an organisational communication network akin to that of a hospital or other large institution, was used to run the simulation framework. In order to distinguish between communications within and between departments, the dataset included synthetic attributes for sender_team and receiver_team. The baseline communication network was built using a six-month time slice (January 2001 to June 2001) following data cleaning and preprocessing. In order to represent a realistic operational phase where communication patterns tend to stabilise before and after intervention strategies, this period length was selected.

Thousands of weighted edges that represented communication frequency and several hundred nodes (individuals) made up the final baseline network. Although the network was connected, only a small portion of the possible communication links were active, as evidenced by the low network density (ranging between 0.02 and 0.05). With an average degree of roughly 4 to 6, it was confirmed that most people only interacted with members of their own teams. Strong internal silos were evident when comparing intra- and inter-team edges, with nearly 70–80% of all communications taking place within the same department.

With a baseline modularity value ranging from 0.55 to 0.65, the Louvain community detection algorithm identified several dense clusters that corresponded to team boundaries. The network structure was highly compartmentalised, with communication mostly restricted to a small number of close-knit groups with few intergroup connections, as indicated by the high modularity score. These findings are consistent with actual situations seen in healthcare institutions, where specialised departments such as the intensive care unit, radiology, and administration effectively communicate within their own units but find it difficult to maintain cross-functional coordination.

The simulation of communication interventions was conducted after the baseline network characteristics were determined. Cross-team liaisons, weekly cross-team rituals, and rotation policies were the three types of interventions that were put into place one after the other and assessed separately. To keep experimental conditions under control, each intervention was applied to a duplicate of the baseline graph. The change in modularity (ΔQ), network density, and inter-team edge ratio following the intervention were used to gauge how effective each strategy was.

The biggest improvement was obtained from the cross-team liaison intervention. Modularity decreased from an average of 0.58 to 0.46, indicating a ΔQ of –0.12, by fictitiously adding direct links between members of various departments. This significant decrease suggests that silo barriers are weakened by the introduction of structured cross-departmental communication roles, such as integrated project teams or coordination officers, which successfully connect disparate clusters. Furthermore, the network density exhibited a slight

positive shift, suggesting greater connectivity without oversaturation, and the inter-team communication ratio rose by almost 15%.

Modest modularity was reduced (from 0.58 to 0.50, $\Delta Q \approx -0.08$) by the weekly cross-team ritual simulation, which randomly raised the likelihood of inter-team interactions. Even though the improvement was not as significant as it was in the liaison model, it still showed that regular, consistent cross-departmental meetings or cooperative check-ins help to keep the organization's communication balance. This intervention's probabilistic design resulted in dispersed but weaker links, which are typical of "small-world" networks where information can spread swiftly even through sporadic connections.

Modularity was reduced the least by the rotation policy intervention, which simulated a temporary reassignment of 5% of employees across teams (from 0.58 to 0.54, $\Delta Q \approx -0.04$). Even though this change was only slightly noticeable in the short term, the pattern indicates that as employees integrate networks and knowledge from prior departments, the strategy may have larger long-term benefits. Indicating that reassigned individuals became active connectors within their new groups, the rotation policy also resulted in a slight increase in average degree.

The quantitative results were further validated by visualising the communication networks. The presence of silos was visually confirmed in the baseline graph by the distinct clusters and the close grouping of nodes of the same colour, which indicates community. The visual network structure showed more overlapping clusters and increased interconnectivity following the implementation of the cross-team liaison intervention, indicating improved cross-functional communication. While the rotation policy graphs showed a subtle diffusion of departmental boundaries, the weekly rituals simulation displayed similar trends, albeit with more dispersed cross-links.

In contrast to probabilistic or behavioural interventions, structured interventions—more especially, cross-team liaisons—achieve a stronger and faster reduction in silo strength, according to the comparative analysis of all interventions. This result is in line with the body of research on organisational network modification, which contends that formal role-based interventions have a greater effect since they alter the network's structural characteristics directly. On the other hand, cultural or behavioural shifts like rotations or rituals typically have a slow and indirect impact on communication.

The findings demonstrate that, from a managerial standpoint, dismantling dynamic silos necessitates deliberate network design rather than just promoting unofficial cooperation. In a hospital setting, this can entail creating digital communication channels that guarantee fair access to information, hiring specific personnel to serve as liaisons between departments (such as ICU–Radiology or Administration–Emergency), or putting in place structured inter-departmental reviews. The results of the simulation offer measurable proof that decision-makers should give priority to interventions that produce quantifiable gains in teamwork with the least amount of disturbance.

In conclusion, the findings show that the suggested simulation model accurately measures how various organisational interventions affect the effectiveness of communication. The empirical results support the hypothesis that specific network modification techniques, particularly liaison-based strategies, can improve interdepartmental connections, drastically reduce modularity, and turn siloed organisations into integrated communication ecosystems. These revelations offer a solid basis for applying the model to actual hospital datasets, allowing for data-driven tactics to enhance organisational resilience and maximise information flow.

5. Conclusion

In order to model and examine communication silos in complex organisations, this study offers a simulation-driven methodology. By combining graph theory, community detection, and organizational network simulation, the framework enables quantifiable evaluation of intervention strategies without needing real-world disruption.

The best way to break dynamic silos, according to the results, is to implement structured cross-team liaisons. Regular interdepartmental rituals and rotational policies come next. The model offers a scalable template that can be immediately implemented in government organisations, corporate settings, or hospital communication systems with the goal of increasing coordination effectiveness. Future projects will consist of incorporating actual hospital communication datasets, such as chat, email, or incident logs, extending simulation for real-time monitoring to temporal multilayer networks utilising machine learning to make dynamic predictions about new silos. This framework extends our knowledge of how communication interventions can build resilient, cooperative, and adaptable organisations by bridging the gap between simulation-based experimentation and empirical network analysis.

References

1. Waal, A., Berg, R. (2025). Increasing Collaboration Through Silo busting: A Case Study. Schmalenbach Journal of Business Research. DOI:10.1007/s41471-025-00223-8
2. Bunger, A. C., Yousefi-Nooraie, R., Warren, K., Cao, Q., Dadgostar, P., Bustos, T. E. (2023). Developing a typology of network alteration strategies for implementation: a scoping review and iterative synthesis. Implementation Science, 18(10), DOI:10.1186/s13012-023-01266-3
3. Yang, D., Holtz, D., Jaffe, S., Suri, S., Sinha, S., Weston, J., Joyce, C., Shah, N., Sherman, K., Hecht, B., & Teevan, J. (2021). Dynamic Silos: Increased Modularity in IntraOrganizational Communication Networks during the Covid-19 Pandemic. arXiv preprint arXiv:2104.00641, https://doi.org/10.48550/arXiv.2104.00641
4. Everton, S. F., Carley, K. M., & Reddy, M. (2016). Undermining and Strengthening Social Networks through Network Modification. NeurIPS Workshop on Networks in the Social and Information Sciences. arXiv:1602.06461
5. Nugroho, Y., & Fitriani, D. (2023). The Evolution of Organizational Design: The Case of Digital Startups. Qualitative Economics & Management Studies, 2(1), 45–63. DOI.org/10.3982/QE809
6. Saka-Helmhout, A., Geppert, M., & Becker-Ritterspach, F. (2022). Hierarchy and diffusion of organizational forms. Frontiers in Psychology, 13, 932273. DOI:10.3389/fpsyg.2022.932273
7. Patterson, P. D., Pfeiffer, A. J., Weaver, M. D., Krackhardt, D., Arnold, R. M., Yealy, D. M., Lave, J. R. (2013). Network analysis of team communication in a busy emergency department. BMC Health Services Research, 13(109), 1–12. https://doi.org/10.1186/1472-6963-13-109
8. Oldham, M., & Daubner-Siva, D. (2023). Changes in formal structure towards selfmanaging organization and their effects on the intra-organizational communication network. Journal of Organization Design, 12, 1–21. DOI:10.1007/s41469-023-00143-z
9. Wu, Y., Zhu, Q., & Wu, C. (2023). How Organizational Agility Promotes Digital Transformation: An Empirical Study in China's Public Sector. Sustainability, 15(14), 11304. DOI:10.3390/su151411304
10. Ogilvie, C., Ospina, J., Konstantinou, C., Vu, T., Stanovich, M., Schoder, K., & Steurer, M. (2020). Modeling Communication Networks in a Real-Time Simulation Environment for Evaluating Controls of Shipboard Power Systems. arXiv preprint arXiv:2008.10441.

Recent Advances in Hospital Information Systems: Towards Smarter Healthcare

Prarthana M, Sivakumar R

Vellore Institute of Technology, Vellore, Tamil Nadu, India
prarthana.m2022@vitstudent.ac.in, rsivakumar@vit.ac.in

Abstract. Hospital Information Systems (HIS) have recently advanced exceptionally well into ecosystems that enhance patient care, streamline operations and data-driven decisions. Recent inclinations involve the use of Artificial intelligence (AI) and predictive analysis for assistance and early diagnosis, cloud-based platforms and blockchain technologies. Real time applications for monitoring and alerting in healthcare can be significant. In spite of the advantages, the criticisms can be huge such as the high maintenance costs, reliability issues, workforce unwilling to be flexible with adapting to these trends, data privacy concerns and cybersecurity risks etc. are serious concerns taken into matter. The resolution of these problems is essential in achieving a balanced approach that combines innovation with robust governance for a future where hospital information systems are seamless, patient-centric, and capable of delivering measurable improvements in both clinical outcomes and operational efficiency.

Keywords. HIS, artificial intelligence, predictive analysis.

1. Introduction

The medical field has experienced a radical change in the last twenty years due to the increased usage of Hospital Information Systems (HIS) - the combined platforms that are aimed at controlling, storing, and sharing the patient-related information efficiently. A Hospital Information System is the digital core of the contemporary hospitals and promotes clinical, administrative and financial tasks. The main goal is to simplify the workflow, improve patient care, and make sure that the correct information is easily available to the health care workers at all levels of the service provision. In the past, healthcare data management was very dependent on manual record keeping and paper records. Not only did this traditional approach add much administrative weight, it also added delays, data duplication and loss of important information. The advent of the Electronic Medical Records (EMR) and Electronic Health Records (EHR) in medical informatics has brought a revolution in making the process of gathering and accessing patient history, lab results, prescriptions, and discharge brief easier by digitalizing the whole process. All these systems together constituted the basis of data-driven decision making in health care institutions.

2. Literature review

2.1. Recent Developments in Clinical Text Summarization

More recent work emphasizes the increasing need for effective and high-quality summarization methods for the healthcare industry, where electronic medical records (EMRs) and clinical notes tend to be lengthy, technical, and complex. The application of domain-based knowledge structures, such as the PICO model (Patient, Intervention, Comparison, Outcome), has been crucial in improving summarization precision and clinical significance. The DKGE-

PEGASUS model is a prime example of this style by incorporating PICO-guided domain features through graph convolutional neural networks, thus enhancing factual informativeness and consistency over general-purpose models such as BERTSum, T5, and KeBioSum. Empirical comparisons on PubMed and CDSR corpora show superior performance of domain-guided summarizers, highlighting their promise for evidence-based clinical practice and decision support at warp speed [4][3][5].

Concurrent improvements in Retrieval-Augmented Generation (RAG) and Generative AI have revolutionized the landscape of automated summarization. The LLAMA + RAG model exhibits outstanding ability to generate structured summaries and identify risk factors from diverse clinical data. RAG's retrieval layer tames hallucination errors characteristic of zero-shot LLMs, resulting in summaries that are contextually plausible and factually accurate. Validation across aged-care datasets achieves rates of accuracy nearing 99%, aligning almost perfectly with expert-manufactured gold standards. These models are especially useful in transforming unstructured EHRs into actionable medical information, thereby enabling timely diagnosis and intervention [3][1].

These are supplemented by Clinical Concept-Guided Transformer Models that incorporate medical ontologies like SNOMED-CT to direct extractive and abstractive summarization. These models improve relevance and completeness by concentrating on chief clinical issues, interventions, and outcomes in discharge summaries and progress notes. Ensemble architectures that merge top-k extractive ranking and abstractive generation have proven to be better at generating clear, actionable outputs while handling the issues with long and multi-document inputs [5]. However, challenges like factual coherence, balance between extractive and abstractive, and human-in-the-loop validation are still areas requiring ongoing improvement.

2.2. Clinical Coding Data Interpretation with Summarization Assistance

The application of automated clinical coding, including the allocation of standardized codes like the International Classification of Diseases (ICD-10) to free-form clinical text, continues to be a challenging task in health informatics. In this area, the use of text summarization as a key and facilitatory technology is increasingly important. Conventional multi-label classification models, commonly used as they are in the case of lengthy clinical reports, commonly fail. These models face a dual challenge: an overwhelmingly high-dimensional label space, with thousands of possible ICD codes, and the redundant, frequently noisy, nature of the text input. This combination degrades performance, especially in detecting subtle or complicated diagnoses.

To correct for these particular shortfalls, the SuM framework has been suggested as a new architecture that optimally incorporates summarization into the coding process [2]. This architecture functions by initially extracting the most critical diagnostic and procedural details from the original text. This is done through knowledge-guided pointer networks, which are designed to pick out and extract the most vital information in the first place, thereby eliminating the redundancy and noise of the original note. After distillation, the architecture utilizes hierarchical attention mechanisms to align the aggregated summary intelligently with the official, structured ICD code descriptions. This alignment process is essential for projecting the distilled clinical facts onto their associated formal codes. The performance of this strategy is shown in comparative benchmarks, where the SuM approach makes dramatic gains in precision and recall. These gains are especially significant for long-tailed and rare code distributions, which are notoriously challenging for generalized models. This performance places particular emphasis on the importance of summarization towards making not just the

automated coding process more efficient, but fundamentally increasing the representational fidelity of the output clinical data [2].

Nonetheless, notwithstanding such specialist advances, there remain deep-seated limitations across the existing technology landscape of clinical summarization and generation systems. A key and persistent problem is overdependence on small, domain-specific datasets. Most state-of-the-art models are trained on individual-institution training sets, which presents a serious risk of sampling bias. These models will likely not generalize when utilized in other healthcare institutions with disparate patient populations, documentation habits, or case mixes. Adding to the problem is the ongoing threat of "hallucination" in generative output. Abstractive models can generate text that is fluent and plausible but factually inaccurate or unsupported by the source document, a serious failure in the high-stakes clinical setting.

Meeting these daunting challenges necessitates an interdisciplinary approach to enhance model robustness, reliability, and generalizability [1][4]. A key step is the incorporation of multi-domain ontologies, which can endow models with a richer, standardized medical knowledge base that is beyond any one institution. In addition, the area needs to be directed towards the creation and application of multi-centric large-scale clinical corpora in order to make sure that models are trained on a more representative and diverse set of clinical information. Lastly, this evolution needs to come with the creation of solid evaluation protocols. These protocols need to go beyond mere fluency measures to strongly evaluate the factuality and clinical accuracy of generated summaries, making these systems safe, accurate, and effective for deployment in various healthcare environments [3][5][1][4].

2.3. Large Language Models for Medical Summarization

Recent advances in large language models (LLMs) have transformed clinical summarization tasks. The QLoRA (Quantized Low-Rank Adapter) fine-tuning method, used for LLaMA 2 models, allows for memory-efficient adaptation in hospital discharge summaries [9]. The comparative study suggests that QLoRA-tuned models perform better than prefix-tuning and null-prompting baselines, scoring higher BLEU and ROUGE-L values and clinician preference ratings. The innovation reflects the success of lightweight fine-tuning for domain adaptation without sacrificing privacy—a critical issue in healthcare NLP. Further, QLoRA's performance in preserving high generalization and medical understanding makes it a viable method for real-world EHR summarization pipelines.

Alongside, Ensemble Data Summarization for Medical Diagnosis presents a non-textual approach to summarization through the compactification of large medical data sets for effective anomaly detection [8]. The system integrates clustering algorithms (X-means, GK-means) and bootstrapping, maximizing sample selection through an adapted Chernoff bound. Tests on benchmark datasets including Breast Cancer Wisconsin and Colon Cancer prove to be better in both computational speed and rare anomaly retention. The method greatly improves the efficiency of anomaly detection algorithms such as k-NN, LOF, and CBLOF, highlighting the importance of data summarization outside text in diagnostic intelligence.

2.4. Hybrid and Contrastive Learning Methods in Medical Summarization

Hybrid models are a significant step forward in summarization design. A great case in point is the Hybrid Solution for Summarizing Various Medical Texts, where BioBERT-based extractive selection is paired with a T5-based abstractive generator [12]. This two-stage approach supports factual grounding and fluency, surpassing solely abstractive baselines through the generation of clinically useful, coherent summaries for medical practitioners. Another key contribution is the ChatGPT-based Contrastive Learning (CCL) approach for radiology report summarization [6]. The approach uses ChatGPT to create negative samples

and augment minority cases, reducing data and exposure bias. Experimental evaluation on MIMIC-CXR datasets shows large gains in informativeness, fluency, and factual accuracy, affirmed by expert radiologists. The method is an interdigitating of generative augmentation and contrastive learning, making significant strides in the automatic impression generation of radiological findings. The hybridization trend goes beyond healthcare. In A Production-Grade Hybrid Summarization System for Meetings, a two-stage T5-BART pipeline is introduced, wherein T5 produces an early abstractive draft fine-tuned by BART for linguistic and contextual refinement. Testing on the AMI Meeting Corpus reveals significant improvements over single-model baselines, indicating that ensemble summarization techniques are generally effective across domains [14].

2.5. Emerging Applications and Domain-Specific Adaptations

Aside from healthcare, summarization techniques have been extended to include domain-aware and context-sensitive tasks. In user-oriented situations, the DOMINER framework for Domain Feature Mining utilizes bond energy clustering to recover feature sets from opinion datasets with state-of-the-art precision, recall, and F-scores over traditional methods like Association Rule Mining and Product Aspect Ranking [7]. While intended for user review mining, its scalable approach lays the groundwork for domain adaptation in biomedical summarization, where feature interpretability and relevance are critical.In addition, research comparing ScholarGPT and ChatGPT for academic summarization purposes in anatomy education indicates that both platforms are decent at simple tasks but decline as the complexity of queries increases. ScholarGPT only slightly beats ChatGPT in accuracy and retrieval precision; however, both suffer from citation constraints and pro-positive bias [10]. The above findings emphasize the need to embed AI systems into confirmed academic databases for reliable summarization.

Through the examined literature, a distinct path unfolds—from early repetitive and Bi-LSTM designs to domain-adapted hybrid transformer models and memory-scarce LLM variants. The meeting of retrieval-augmented generation, domain-informed embeddings, and hybrid learning architectures marks a paradigm shift towards contextually fluent, factually accurate, and mission-designed summarization systems. In medicine, these advancements hold out for revolutionary enhancement of clinical documentation, coding automation, and patient understanding, as well as carryover applicability to areas like education and business intelligence. Future studies must continue to tackle challenges of scalability, factuality, and moral deployment in order to realize the potential of intelligent summarization systems in practical applications.

The rapid growth of online text has generated tremendous demand for automated systems that can summarize large volumes of information into summary form, condensed, coherent, and correct. In the past few years, there has been increased emphasis on advances in Natural Language Processing (NLP) with respect to the development of text summarization and generation methods. There has been a definite upgrade from early deep learning architectures to highly advanced, domain-specialized hybrid transformer models. This is reflective of a trend towards systems that not just summarize but also tailor content for individual fields, like healthcare and business, to enhance understanding and productivity. Early deep learning methods served as the grounding for abstractive summarization. One typical approach in founding methods, a research study suggests an abstractive summarizer that employs a sequence-to-sequence model. This architecture makes specific use of Bidirectional Long Short-Term Memory (Bi-LSTM) supplemented by a double attention pointer network. It was formulated to generate accurate and redundant-free summaries, accurately measured using the ROUGE metric on the CNN Daily Mail corpus [11]. Along

with the advancement of the field, the architectures have since been modified to highly specialized applications outside of summarization. For example, into text generation, a paper presents an "Adaptive-Bert Network" for the particular application of generating advertisement text. This framework combines layout constraints uniquely into the generation process. A prediction-evaluation approach initially decides the best word length and font size for a particular layout, and then the modified BERT model produces high-quality text from source blogs that meets these exact spatial constraints [15].

One of the main motivations for current text summarization work is the healthcare industry, where information overload can hinder both professional decision-making as well as patient comprehension. One of the studies tackles the issue of making complicated medical reports understandable to patients. It suggests a system based on Retrieval-Augmented Generation (RAG) and fine-tuned versions of the LLAMA models for both structured documents and unstructured patient-reported text. The aim is to improve patient understanding by providing summaries that are not only short but also contextually relevant and comprehensible and, in the process, enhancing the patient-provider conversation [13]. In parallel, other work aims at healthcare professionals with the creation of new solutions to summarize various medical texts. One of these presents a hybrid approach that begins with an extractive stage, driven by a domain-specific knowledge embedder (BioBERT), in order to extract clinically important content. This is subsequently followed by an abstractive stage through a T5 transformer so that the ultimate summary is coherent and readable. This fusion of domain expertise is demonstrated to substantially enhance the quality and applicability of summaries for use in clinical practice [12].

To address the limitations of individual models, scientists are now creating advanced hybrid systems by fusing the capabilities of multiple architectures. This is especially the case for extremely complex, domain-specific tasks. For the challenging domain of multi-speaker meeting summarization, a production-level system is introduced that marries two strong transformer models. This two-stage pipeline initially employs T5 (Text-to-Text Transfer Transformer) to create a first-draft abstractive summary. The draft is then polished by BART (Bidirectional and Auto-Regressive Transformer) to improve linguistic fluency, contextual saturation, and discourse coherence. This hybrid model clearly surpasses single-model baselines on the AMI Meeting Corpus [14]. The medical summarizer described above, which combines a domain-specific extractive module (BioBERT) with an abstractive generator (T5), is another instance of this trend. This approach guarantees the produced summary is based on the most significant clinical facts found by the extractive step, demonstrating the strength of hybrid, domain-aware models.

3. Methodology

The suggested research method centers around planning and executing an AI-based Hospital Information System improvement framework based on Generative AI that will increase the unstructured clinical text into structured and interpretable information. It is based on natural language processing (NLP) and deep learning architectures, which automate 2 essential phases of hospital data management: departmental classification and summarizing clinical notes. The methodology was developed to overcome three major issues uncovered in existing HIS ecosystems - the challenge of the unstructured textual data processing, the lack of context-based cognition in older systems, and the inefficiency of the manual data routing. The system can learn both linguistic and contextual accuracy by combining the transformer based systems like FLAN-T5 in generative summarization and BERT based systems in discriminative classification. The general workflow is a pipeline consisting of a series of steps, starting with the data preprocessing, proceeding through model selection and fine-tuning, and

finally the optimization of the output by a confidence threshold, and performance appraisal by the common NLP performance measures. The models are trained and tested on clinical datasets that are publicly available and artificially generated hospital notes to be diverse and strong. The approach focuses on the domain adaptation and responsible AI deployment so that the model predictions can be deemed as clinically reliable. Besides, low-confidence outputs should still human monitored, in order to maintain transparency and ethical adherence, and the system is not only intelligent, but also trusted in the practice of healthcare.

3.1. System Overview

The suggested model incorporates Generative Artificial Intelligence (GenAI) in the information systems of hospitals to automate two important procedures: (i) clinical text summarization and (ii) classification of medical specialties. The system is made as a modular pipeline that can accept any unstructured electronic medical records (EMRs) or clinical notes and convert them into compressed representations that are short and structured and that can facilitate healthcare workers to make decisions more quickly and direct departments in their routing. Its fundamental design is based on four functions stages which include data gathering and pre-processing, generative summarizing, deep learning classification and visualization of its output which are all part of the end-to-end automation of the entire system. This starts when EMR data is ingested and can be in form of physician notes, discharge summaries or even diagnostic reports. Given that the length of such data can be long, redundant and that they may have different linguistic structures, a large preprocessing pipeline is deployed to normalize the input text. The resulting text is subsequently introduced to the Generative Summarization Module, an implementation of a fine-tuned FLAN-T5 model, a version of the Text-to-Text Transfer Transformer created by Google, selected due to its ability to generate few-shot texts and its ability to generate contextually consistent words. The step of summarization is used to make sure that long medical notes are summarized into clinically meaningful and legible briefs without the critical details of the diagnostic or treatment being missed. After summarization, the processed output is sent to the Department Classification Module, which is a transformer-based deep learning classifier (developed around BERT architecture), which recognizes the most relevant medical specialty, or hospital department relevant to the condition of the patient. The classification can be used to directly direct records to the relevant medical team (e.g., cardiology, neurology, oncology) so that the time spent sorting the data manually is removed in hospital data processes. This is done by using Flask to use a web-based user interface to facilitate easy interaction between the system and end users who might be part of the medical staff, students, or data analysts. The system takes clinical text inputs, real-time summarizes and classifies and presents the two outputs in a structured and understandable format. Also it could be combined with architecture-based on cloud deployment, and it can be scaled horizontally and be integrated with the existing hospital information systems through RESTful APIs. In general, the offered methodology is expected to cover the gap between unstructured medical information and practical clinical analysis. Generative AI use promotes the interpretability and utility of HIS data, and the deep learning classifiers promote intelligent triage and inter-departmental data flow. The framework is modular and interoperable, and therefore can be incorporated into both current and future generation hospital information infrastructures very easily.

3.2. Data Description

The quality and representativeness of underlying data form the major determinant of the performance and reliability of any intelligent hospital information system. The training and evaluation data in this research will be a set of de-identified clinical notes and electronic

medical records (EMRs) obtained through publicly available healthcare repositories and artificially created hospital data. It is a mixture of both actual and simulated records so that there are different medical specialties, and yet, one does not violate ethical and data privacy standards. Every record includes patient demographics, symptoms, medical history, diagnostic impressions, and treatment plans, which makes them a solid foundation of both text summarization and department classification activities. Before the model is trained, the dataset is subjected to an intensive preprocessing pipeline that is meant to eliminate noise and promote textual consistency. The raw EMRs are usually full of non-standard abbreviations, typing mistakes, and format debris (removing) caused by manual entry or optical character recognition (OCR) systems. Such discrepancies are systematically tackled with the help of tokenization, case normalization, and expansion of abbreviations with the help of a medical lexicon. Stopword elimination is selectively used to preserve domain specific terminologies vital in diagnosis interpretation like drug names, medical operations as well as physiological parameters. Furthermore, any kind of personally identifiable information (PII) is substituted with the anonymized placeholders to adhere to the data privacy standards like the HIPAA and ICMR regulations. In order to learn successfully, the data is divided into structured pairs of input text (original clinical note) and output summary (reference or model-generated text). In the same manner, in the classification activity, a record is labeled with a label of the medical department of that record, based on the primary diagnosis or clinical keyword. The sample covers about 5-8 major departments of the hospital such as cardiology, neurology, orthopedics, gastroenterology, oncology and general medicine thus providing a wide coverage of the hospital operations. To avoid model bias, data distribution among departments is equal and stratified sampling in both training and testing is used in the process in order to have representational fairness. Data augmentation methods are used on the textual records in order to increase generalization and strength further. These consist of synonym replacement, paraphrasing of sentences with pretrained transformer models, and small-scale syntactic noise. These extensions artificially approximate the variability of clinical documentation style in the real world setting and enhance the flexibility of the generative model to different hospital settings. Also, ethical aspects are the main focus of the data design. Medical text contains sensitive data, and, thus, a high level of data anonymization, compliance with institutional reviews and transparency in the use of the model are provided in this work. None of the personally identifiable data is stored or exchanged at any point. Interpretability and human control are also given a significant role in system design so that generated summaries and classifications are not independent clinical decisions, but decision-support tools. The suggested system is based on this carefully selected, preprocessed, morally consistent dataset and constitutes a good base towards the application of correct and context-sensitive GenAI models that may be utilized to improve information systems within hospitals with reliability and transparency.

3.3. Model Architecture and Training Strategy

The proposed architecture combines a Generative Ai-based summation model and a deep learning-based medical department classifier on a dual-model basis to achieve results. The two parts work in sequence to convert unstructured and raw electronic medical records (EMRs) into brief and formatted and department-specific results. The structure takes advantage of the strength of already trained transformer models, which have proven to be the best in natural language processing tasks especially in the medical field where the coherence of context and language fluency is the most critical part.

3.3.1 Generative Summarization Module

Implementation of the summarization task is undertaken with the help of FLAN-T5 (Fine-tuned Language Net--Text-to-Text Transfer Transformer), which is a generative transformer model created by Google Research. The FLAN-T5 was chosen because it has outstanding generalization ability in various NLP tasks, its high ability to process long-form text, and interactions to specific instructions when given prompt-based conditioning. The curated corpus of clinical notes is fine-tuned on the model in this study to make sure that it is adjusted to linguistic patterns and terminologies of the medical domain. In the fine-tuning, the model is trained to produce abstractive summaries, i.e., the ones, which paraphrase and summarize information instead of extracting its fragments. The training cases are defined by each being a source EMR input and a reference summary output. The model is trained so as to reduce the loss of cross-entropy between target sequences and predicted sequences. Instruction-tuning prompts, e.g. the summary of the subsequent medical note in a concise way, serve to direct the model in clinically coherent productions. In order to avoid overfitting, dropout regularization and early stopping methods are used. Such evaluation measures as ROUGE and BLEU scores are calculated to evaluate the quality of summaries, coherence and content retention.

3.3.2 Department Classification Module

In the classification part, a Bidirectional Encoder Representations of Transformer (BERT) model is used as a fine-tuned model that carries out text classification tasks. Once the step of summarization is completed, the produced text is transferred to this module to guess which department in the hospital will be the most useful. The reason behind the selection of BERT is the fact that it has a bidirectional attention mechanism that can understand the context of words both before and after each other, a key characteristic of the identification of clinical entities and relationships. The last dense layer of the classifier produces probability distributions among various department labels by use of SoftMax activation. To achieve the stability of convergence, the model is optimized by the Adam optimizer and a learning rate optimizer. In order to increase the ability of the model to comprehend medical language, a domain-specific version, BioClinicalBERT, that has been pre-trained on a large set of de-identified medical notes in the MIMIC-III dataset replaced the generic BERT classifier. In contrast to the general BioBERT model that acquires linguistic representations by using open-domain datasets, like Wikipedia and BooksCorpus, BioClinicalBERT is specifically trained to learn complex medical terms, abbreviations and clinical syntax. This specialization gives it a better accuracy and domain in the interpretation of contextually rich text. This technique is called Domain-Specific Transfer Learning and it avoids catastrophic forgetting of primary medical semantics in the process of better aligning with the seven standardized hospital department labels. Another optimization was done with a confidence-based decision threshold, so that predictions with less confidence (e.g. 70) are marked to be reviewed by humans. This mixed methodology makes it safe and interpretable in clinical practice as it brings machine efficiency and expert control together. Therefore, as the triage system of a hospital, the model will have close to perfect accuracy since any high-risk cases of uncertainty will be sent to humans to be confirmed. Medical domain adaptation was shown to be effective since empirical assessment showed a steady accuracy increase over baseline BERT fine-tuning. In addition, the ClinicalBERT classifier was also found to converge faster and more resistant to noisy input text, highlighting the advantages of training on a domain-aligned model as compared to retraining completely.

3.3.3. Integration and Training Strategy

The two modules are combined into a python pipeline through Hugging Face Transformers. The training and inference are also performed on the environments based on the GPU (NVIDIA RTX 3090) to speed up the computations. The end-to-end system is made to be modular, such that the summarizer and the classifier can be used, either alone or together, in the same workflow. In evaluation, the summarization model results are given to the classifier and the results compared to manually labelled department data. This combined training methodology makes both modules work semantically, which minimises the errors in propagation and enhances their overall accuracy. The proposed model is able to capture medical semantics, provide clinically faithful summaries, and predict department classifications accurately through this hybrid generative-discriminative architecture, which can help build smarter and more autonomous hospital information systems.

3.4. Tools and Implementation Framework

The proposed hospital information system, which is based on GenAI, was implemented by integrating both the latest software frameworks and scalable hardware infrastructure to guarantee robustness, reproducibility and scalability to varying healthcare settings. The Python language was used in the development of the system because of the rich data processing, machine learning, and natural language understanding ecosystem. The workflow architecture is a combination of open-source software like Hugging Face Transformers, PyTorch, and Flask, which can be used collectively to provide an end-to-end system of model training to real-time web deployment.

The model training scenario was set up on an NVIDIA RTX 3090 24GB VRAM and 64 GB system memory, high-performance computing system. PyTorch was the main deep learning platform to fine-tune a model and make inferences due to its dynamic computation graph and the ability to be provided with a strong degree of compatibility with the acceleration of the model on the GPU. The implementation of both the FLAN-T5 summarization model and BERT-based classification model was done via the Hugging Face Transformers API, which allowed both pretrained models and tokenization tools to be easily integrated. The datasets and transformers libraries offered effective data processing pipelines whereas preprocessing, data partitioning, and performance analysis were performed by scikit-learn and NumPy.

To have the frontend- backend interaction, a lightweight web framework, Flask, was used to serve the models as RESTful APIs. The Flask server processes the HTTP POST requests with clinical text input and the parameters sent and provides model-generated outputs, such as the summarized report and the predicted medical department. The frontend interface was created with HTML, CSS and JavaScript which guaranteed a user-friendly user interface fit to the healthcare professionals and research staff. Its modular web architecture allows it to be flexibly deployed to something like a local server or a cloud provider like AWS or Google Cloud, which in turn will provide scalability and can be easily integrated into current information systems of a hospital.

The flow of work in the system starts when the user enters a clinical note with the help of the interface. The Flask backend is the one that sends the text to the preprocessing pipeline, which tokenizes, removes noise, and anonymizes the text. The text is then cleaned and sent to the FLAN-T5 summarization engine that produces a readable and concise version of the text sent to it. The resulting summary is then forwarded to the BERT-based classifier that labels the summary with a departmental label after being trained to learn contextual features. Both of these are then displayed to the user interface in formatted fashion with the confidence scores attached along with optional visualization items like bar plots showing prediction probabilities.

In order to make the project reproducible, maintainable, and structured in a modular manner with a distinct separation of modules, the project is organized into data handling, model inference, and user interaction modules. It also has in-built logging and monitoring tools to monitor the API performance and error management during inference. Moreover, environment encapsulation was developed with Docker containerization that enables the same activity to be run on diverse hardware and software platforms.

The deployment pipeline was given a lot of focus on security and compliance. Every interaction between the client and server is encrypted by HTTPS protocols and the system complies with healthcare data manipulation standards by not storing any identifiers or clinical records of patients after inference. These design choices will provide ethical deployment without violating user privacy and trust. Overall, the implementation model shows that it is possible to build an advanced GenAI-powered hospital information system based on the available and open-source technology and still ensure that clinical integrity, scalability, and usability remain consistent in the healthcare setting.

4. Results

To find out the effectiveness, precision, and dependability of the proposed Generative AI-based hospital information system in processing unstructured clinical data, the experimental appraisal of the system was carried out. The analysis of the results was conducted in both a quantitative way, in terms of predetermined metrics of evaluation, and in a qualitative form, in terms of the relevance and interpretability of the model results to the clinical picture. The performance of the summarization and classification modules was evaluated separately and combined together to guarantee the integrity of the integrated system at its end.

4.1 Summarization Performance

The summarization model based on FLAN-T5 showed good performance when it comes to the production of coherent and contextually accurate summaries of electronic medical records. Assessment has been done based on conventional measures like confidence scores, which are the measures of a similarity of generated and reference summaries. The review of the summaries and readability as well as retention of information must also be confirmed through manual evaluation students and professionals.

A comparative study to baseline extractive algorithms demonstrated that the generative algorithm gave shorter and more human-like summaries. An example would be extractive models, which were more likely to reproduce redundant sentences and was not effective at rephrasing results but FLAN-T5 generated compact but semantically rich representations of medical cases. A test case of a cardiology patient proved that the model was able to reduce a 270-word note to a 90-word summary conserving all vital signs, medication, and diagnosis details with little loss of facts.

4.2 Department Classification Accuracy

The overall BERT-based classification module scored 82.5 and 85.2 in terms of accuracy and F1-score respectively, in eight large hospital departments. The confusion matrix showed that high-volume categories, including Cardiology and Neurology, performed very well with a minor error rate in the departments that have similar terminologies, including General Medicine and Endocrinology. Summarization output integration followed by classification was found to yield better results by up to 6-percent accuracy gain, which validates the claim that compact and uniform summaries provide reduce noise and offer better contextual information to the classifier.

4.3. Discussion and Insights

The results ensure that there is a possibility of the both the generative and the discriminative AI model complementing one another in the context of healthcare. The summarizer enables the interpretability of data whereas the classifier enables efficient triage and automation of workflow. However, there were some shortcomings noticed. The model was sometimes found to be hard to deal with rare diseases and those involving multi-departmental cases where the symptoms overlapped and therefore there was confusion in classification. Also, the model was prone to data imbalance; the performance of the model was lower when the data samples used by specific departments were small.

Regardless of these shortcomings, the presented GenAI-based HIS framework has very specific advantages in the form of accuracy, speed of processing, and clarity of information. The methodology preconditions the use of large language models in the workflow of hospitals by making the provision of decision support quicker and more accurate.

The results confirm that domain specialization has quantifiable benefits that extend beyond predictive performance to include interpretability and safety, which are the key characteristics of vital significance to healthcare AI. The proposed system having FLAN-T5 to perform generative summarization and BioClinicalBERT to perform discriminative classification represents a hybrid type of generalist-specialist architecture that can process a wide and deep range of medical data. This bi-polar paradigm provides a scaled framework in the development of future intelligent hospital information systems and creates gateways to multimodal extensions with the integration of textual, imaging, and diagnostic streams of data.

4.4. Comparative Analysis: General BERT vs. ClinicalBERT

A comparative analysis was conducted to assess how domain-specific transfer learning will affect the output of the proposed department classification module. Though the previous version used a general-purpose BERT model that was trained on open-domain text collections, the new architecture is based on BioClinicalBERT, a transformer that was pre-trained on clinical notes and biomedical literature only. This analogy identifies the power of domain specialization in enhancing classification accuracy, interpretability and reliability in medical information systems

5. Conclusion

This paper presented comprehensive research into the recent trends on the integration of hospital information systems through Generative Artificial Intelligence (GenAI) in text summarization of medical documents and classification of departments. The proposed dual-stage model has been useful in overcoming one of the oldest problems in healthcare informatics, which is the interpretation of unstructured clinical narratives. The system had a remarkable improvement in accuracy and interpretability compared to traditional methods with the integration of the FLAN-T5 model of brief and contextually accurate clinical summary and domain-specific transfer learning utilizing BioClinicalBERT that could classify into a specific department. Operational reliability was also offered by the combination of the mechanism of confidence-threshold, which allowed human review of questionable predictions and, therefore, complied with the principles of ethical considerations regarding the application of AI to clinical activities. Altogether, the study highlights the possibility of changing the GenAI in order to develop smarter, interoperable, and context-aware hospital information systems. The obtained results underline the importance of using domain-adapted language models to maximize the effectiveness of medical processes, reduce administrative burdens, and improve care delivery to patients. In the long run, the integration of such smart systems

into the actual hospital environments can open a new dawn of automation in healthcare whereby unstructured data will be automatically used to generate actionable data. This effort can be expanded on in future research with multimodal data streams, explainability AI systems, and real-time deployment systems and in this way progress to an autonomous but ethically responsible digital health ecosystem.

References

1. S. Rahulprasath, A. Lakshithraj, S. Pranav Harshan, and J. Sreemathy, "AI in Healthcare: Simplifying Medical Reports for Enhanced Patient Comprehension," in 2025 International Conference on Emerging Technologies in Computing and Communication (ETCC), 2025, doi:10.1109/ETCC65847.2025.11108371.
2. Y. Sun et al., "Enhanced ICD-10 code assignment of clinical texts: A summarization-based approach," Artif. Intell. Med., vol. 156, p. 102967, 2024, doi: 10.1016/j.artmed.2024.102967.
3. M. Alkhalaf, P. Yu, M. Yin, and C. Deng, "Applying generative AI with retrieval augmented generation to summarize and extract key clinical information from electronic health records," J. Biomed. Inform., vol. 156, p. 104662, 2024, doi: 10.1016/j.jbi.2024.104662.
4. S. Jiang, Q. Zheng, T. Li, and S. Luo, "Clinical research text summarization method based on fusion of domain knowledge," J. Biomed. Inform., vol. 156, p. 104668, 2024, doi:10.1016/j.jbi.2024.104668.
5. T. Searle, Z. Ibrahim, J. Teo, and R. J. B. Dobson, "Discharge summary hospital course summarisation of in patient Electronic Health Record text with clinical concept guided deep pre-trained Transformer models," J. Biomed. Inform., vol. 141, p. 104358, 2023, doi:10.1016/j.jbi.2023.104358.
6. Z. Luo et al., "ChatGPT based contrastive learning for radiology report summarization," Expert Syst. Appl., vol. 267, p. 125827, 2025.
7. H. K. Thakkar, P. Singh, and Y. Kumar, "DOMINER: Domain Feature Mining from Unstructured Data for Effective Text Summarization," in Proc. Int. Conf. Mach. Learn. Data Eng. (ICMLDE 2023), Procedia Comput. Sci., vol. 235, 2024, pp. 559–567.
8. M. Ahmed and A.N.M. B. Rashid, "EDSUCH: A robust ensemble data summarization method for effective medical diagnosis," Digit. Commun. Netw., vol. 10, no. 1, pp. 182–189, 2024.
9. J. Goswami, K. K. Prajapati, A. Saha, and A. K. Saha, "Parameter-efficient fine-tuning large language model approach for hospital discharge paper summarization," Appl. Soft Comput., vol. 157, p. 111531, 2024.
10. D. Chytas et al., "Can artificial intelligence accurately detect and summarize anatomy education literature? A comparative analysis of ChatGPT and ScholarGPT," Morphologie, vol. 109, p. 101061, 2025.
11. Preethi, S., Shibi, K. M. S., Grace, R. K., & Geetha, M. S. (2022). Abstractive Summarizer using Bi-LSTM. 2022 International Conference on Edge Computing and Applications (ICECAA).
12. Samuel, P., Reshmy, A. K., Berna, E., & Sriraj, S. K. (2024). A Hybrid Solution for Summarizing Diverse Medical Texts in the Health Domain. 2024 International Conference on Communication, Computer Sciences and Engineering (IC3SE).
13. S, R., A, L., S, P. H., & J, S. (2025). AI in Healthcare: Simplifying Medical Reports for Enhanced Patient Comprehension. 2025 International Conference on Emerging Technologies in Computing and Communication (ETCC).
14. Kola, J., Pasupuleti, S. P., Kodete, S. P. M., & Kranthi, V. A. (2025). A Production Grade Hybrid Summarization System for Meetings: Combining T5 and BART for Improved Fluency and Contextual Accuracy. 2025 8th International Conference on Computing Methodologies and Communication (ICCMC).
15. Yao, S., Pan, X., Dong, H., Kong, C., & Wu, X. (2023). Adaptive-Bert Network for Advertising Text Generation. 2023 IEEE 6th International Conference on Electronic Information and Communication Technology (ICEICT).

ICU Bed Prediction Model using Ensemble Learning

Adhvaith Shivkumar, Sivakumar Rajagopal

Vellore Institute of Technology, Vellore, Tamil Nadu, India
adhvaith.shivkumar2022@vitstudent.ac.in, rsivakumar@vit.ac.in

Abstract. Accurate forecasting of Intensive Care Unit (ICU) bed occupancy is a critical component of hospital capacity planning and emergency preparedness. This study proposes a data-driven forecasting framework employing advanced ensemble learning techniques to predict short-term ICU utilization based on routinely collected hospital operational data. Temporal dependencies within the dataset were modeled through the incorporation of lag features, rolling statistical measures, exponentially weighted averages, and weekday encoding. The predictive framework integrates three gradient-boosting algorithms—Light Gradient Boosting Machine (LightGBM), Extreme Gradient Boosting (XGBoost), and Categorical Boosting (CatBoost)—combined through a Stacking Regressor with Linear Regression as the meta-learner. Experimental evaluation on the Michigan Hospital Capacity Tracking dataset demonstrates that the proposed model achieves a Mean Absolute Error (MAE) of 37.72 and a Root Mean Squared Error (RMSE) of 60.34, representing an improvement of approximately 20% over baseline persistence models. The ensemble's ability to efficiently capture nonlinear temporal patterns without reliance on patient-level electronic health records highlights its applicability for real-world hospital management systems. The proposed methodology offers a scalable and interpretable solution capable of supporting proactive decision-making in resource-constrained healthcare environments.

Keywords. ICU Capacity Prediction; Ensemble Machine Learning; Time Series Modeling; Healthcare Resource Management; Predictive Analytics

1. Introduction

Management of Intensive Care Unit (ICU) resources is a cornerstone to resilience and responsiveness of healthcare systems. Imbalances in demand and capacity in the ICU may result in critical bottlenecks in patient care during times of high patient influx, such as pandemics, seasonal disease increase, or mass emergencies. Hospitals have traditionally used either manual scheduling or heuristic planning or a mere statistical extrapolation in capacity forecasting. These techniques can be, however, less flexible and less accurate to handle the quickly changing demand trends in the dynamic healthcare environment.

The latest developments in the field of machine learning (ML) have allowed the implementation of data-driven solutions to predict the trends of occupancy in hospitals, staffing management, and resource use. Although the use of techniques like Autoregressive Integrated Moving Average (ARIMA) and Long Short-Term Memory (LSTM) networks in time series modeling have been used in the healthcare sector, they are either constrained by the assumption of linearity (ARIMA) or through high computation costs and complexity in tuning (LSTM). Moreover, numerous previous studies rely strongly on granular electronic health record (EHR) data which is not always inter-institutional and real-time.

To overcome these problems, this paper suggests a gradient boosting-based ensemble model to predict short-term ICU bed capacity based on regularly gathered operational data of the hospital. The presented model combines Light Gradient Boosting Machine (LightGBM),

Extreme Gradient Boosting (XGBoost), and Categorical Boosting (CatBoost) models in a stacked ensemble framework, which is simultaneously optimized by the help of a Linear Regression meta-learner. The structure enables the model to support nonlinear temporal dependence, efficient management of missing or categorical data, and trade-off accuracy and computational efficiency.

The model was improved by using temporal feature engineering methods such as lag variables, rolling averages, exponentially weighted means, and weekday encoding so that it was more able to learn short-term occupancy patterns. The system was tested on the Michigan Hospital Capacity Tracking dataset that is a realistic operating environment in a hospital.

The suggested structure has a better predictive performance than the classical baselines with significant decreases in Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE). The interpretability, scalability, and low computational cost of it make it an appropriate choice in real-time integration into hospital management dashboards, allowing predictive resource planning and evidence-based operational decision-making.

2. Problem Definition and Research Scope

The main aim of this research is to come up with a predictive modeling system that will be able to predict the short-run bed occupancy of the Intensive Care Unit (ICU) in a hospital based on the regularly gathered operational data in the hospital. This is in contrast to the patient-level prediction systems, which are based on specific clinical parameters, and instead, the indicators are at the aggregate level (i.e. ICU use, inpatient hospital admissions, and ventilator usage) that are easily accessible at most healthcare institutions.

The formulation of the problem is as a supervised regression task with the target variable looking at the number of occupied ICU beds on a particular day, predictors being the past measures of utilization as well as the calculated temporal measures. The system is intended to forecast the occupancy of the ICU a few days in advance to help hospital administrators to take proactive decisions regarding staffing, admissions, and allocation of equipment.

In order to do that, the ensemble of the machine learning models consisting of Light Gradient Boosting Machine (LightGBM), Extreme Gradient Boosting (XGBoost), and Categorical Boosting (CatBoost) was created and the stacked generalization approach was used to optimize it. Such a set up enables each of these models to identify different things about the data distribution, such as LightGBM to effectively model large, tabular data; XGBoost to effectively minimize errors using regularized boosting; and CatBoost to effectively treat categorical data and avoid overfitting. The combination of these predictions is a Linear Regression meta-learner that gives the final forecast.

Temporal feature engineering is also included in the framework to model the dependencies on the hospital utilization data. The lag features record short term memory effects, the rolling averages reduce random variations, the exponentially weighted means are more focused on the recent trends, and the weekday encoding is to explain the repetitive nature of operations of a company.

The study has a narrow scope of predicting ICU occupancy only using operational datasets without the use of other covariates including epidemiological data or patient-level data. This controlled limitation indicates the validity of the use of a lightweight, interpretable model in practical hospital information systems without any extra data integration overheads.

3. Current Trends in Hospital Capacity Forecasting

The recent development of analytics in healthcare has been observed to turn towards more of a shift towards data-centered and forecasting model solutions to hospital capacity forecasting. Traditional statistical models such as linear regression and autoregression have been slowly phased out or improved with state-of-the-art machine learning and deep learning models capable of characterizing nonlinear and multivariate relationships in hospital operations data.

It has been studied in a number of works that apply Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) to predict patient admission and bed occupancy using their ability to embrace long-term temporal dependencies. However, neural architecture typically requires enormous datasets, massive computational power, and parameter optimization, and thus are not as practical in real time in resource constrained healthcare environments. Alternatively, hybrid methods that make use of autoregressive models and machine learning algorithms have been proposed as well to increase interpretability and adaptability.

More recently the focus has shifted to ensemble-based models (especially Gradient Boosting Machines (GBM), Extreme Gradient Boosting (XGBoost), and Light Gradient Boosting Machine (LightGBM)) as being more competent with structured time-series data, particularly by virtue of being more competent. Advantages of these models are that they can train faster; they are able to automatically deal with missing values and they have a high level of generalization. Their ability to incorporate both time-related and categorical characteristics has predetermined their suitability to short-horizon hospital occupancy forecasting, employee scheduling, and supply chain management.

In general, the newly developed body of research suggests the growing interest in scalable, interpretable, and computationally effective solutions, which can be easily integrated into hospital management systems to facilitate the prediction of decisions.

4. Model Architecture

The presented ICU occupancy prognostication system is designed in a structured architecture, which integrates data preprocessing, temporal feature engineering, model training, and ensemble-driven prediction. The framework is designed to derive temporal dependencies, nonlinear correlations and operational patterns of hospital utilization data by a sequence of algorithmic and statistical transformations.

The data used is the Michigan Hospital Capacity Tracking System, which includes the daily operational data, which includes the use of the ICU bed, the amount of inpatients, and the use of ventilators. Every record is the status of a day in hospital on a day-by-day basis. In order to make everything consistent, all numerical features were transformed into floating point format and missing values were interpolated with forward filling keeping the continuity in each feature time series and maintaining the temporal order.

The dataset was then cleaned followed by temporal feature engineering to make the data predictively rich. There were three major feature categories that were obtained:

Lag Features: These are the occupancy of the past days so as to encode the short-term memory of the model. As a mathematical expression, to have a target variable y_t (ICU beds used on day t), the lag features may be defined as:

$$y_t-k=ICU_Used\ at\ day\ (t-k), k=1,2,\ldots,7 \quad (1)$$

Rolling Averages: To smooth noise and capture medium-term trends rolling standard deviation and rolling mean were calculated over fixed windows:

$$roll_n(t) = 1/n' \sum_{t=0}^{n-1} y \quad (2)$$

To conduct the present study, rolling mean functions were calculated on both three and seven days windows to reflect short-term and weekly changes in the use of the ICU. These characteristics were referred to as three-day and seven-day rolling averages of the ICU occupancy.3. Exponentially Weighted Mean (EWM): Exponential smoothing was used to help it be sensitive to new changes:

$$EWM(t)=\alpha yt+(1-\alpha)EWM(t-1) \quad (3)$$

where α = 2/ (1 + span), and span was implemented to be 3.

Cyclical Encoding: To capture periodic patterns of workflow in the hospitals, weekday encoding was included. *Pandas* was used to extract the day of the week (0-6) in the date column through the datetime library.

Following the construction of the features, the input matrix X included the lag features, rolling features and statistical features, whereas the target variable y was the real ICU occupancy (ICUUsed). To ensure data leakage was avoided, the data was dichotomised chronologically in a ratio of 80: 20 and shuffle=False in the scikit-learn train-test split function.

The predictive framework uses a stacked ensemble model comprising of three gradient boosting algorithms including Light Gradient Boosting Machine (LightGBM), Extreme Gradient Boosting (XGBoost), and Categorical Boosting (CatBoost) and Linear Regression meta-learner. Every base learner is modeled as an ensemble of decision trees fm(x) with each tree being trained to minimize a differentiable loss function L(y, y) by stage-wise additive training:

$$\hat{y}_t = \sum^{M=1} \quad \nu f(x_t) \quad (4)$$

where m is the learning rate.

Trees in LightGBM are grown in the leaf-wise mode as opposed to the level-wise mode. In every iteration, the algorithm picks the leaf that minimizes the loss the most to be split and the algorithm can converge faster and decrease the error. A split gain is calculated as follows:

$$Gain=GL^2/(HL+\lambda)+ GR^2/(HR+\lambda)- (GL+GR)^2/(HL+HR+\lambda) - \gamma \quad (5)$$

In which G and H are the first and second-order gradient of the loss, l is the L2 regularization term, and g is used to enforce averse splits between leaves.

XGBoost is a supplement to this, building level-wise trees where the loss minimization uses second-order Taylor approximation, which provides efficient regularization of shrinkage and subsampling. CatBoost, however, employs ordered boosting and statistical target encoding to effectively deal with categorical variables without overfitting/bias.

A meta-learning layer combines the predictions of the base learners in a Linear Regression model which determines the best coefficients wi to combine the base outputs:

$$y^=\sum^{i3=1} wi*yi+b \quad (6)$$

The scikit-learn Stacking Regressor was applied in implementation to combine the three underlying models with the meta-learner in a manner that ensures consistency of cross-validation and minimizes overfitting. The last model was trained with 1000 estimators per base learner and 0.05 learning rate. The measures of evaluation, Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) measure the difference between actual and predicted occupancy values:

$$MAE = (1/n)\,\Sigma|y_i - \hat{y}_i|, \quad RMSE = \sqrt{((1/n)\,\Sigma(y_i - \hat{y}_i)^2)} \tag{7}$$

A low value on these measures implies greater predictive accuracy. Lastly, the predictions of the model that had been trained were plotted using an Actual vs Predicted plot, which allowed a clear comparison between actual occupancy trend and the model predictions. This tight fit of these curves confirmed the learning capability of the model to short-term fluctuations but general stability. In general, this architecture is a hybrid, interpretable, and computationally efficiency design, which manages to combine statistical feature extraction and current ensemble learning to forecast ICU occupancy using operational datasets.

5. Results and Model Evaluation

The developed ensemble regression framework, integrating LightGBM (Light Gradient Boosting Machine), XGBoost (Extreme Gradient Boosting), and CatBoost (Categorical Boosting), was trained and evaluated on the Michigan Hospital Tracking dataset. The dataset contained time-stamped information on hospital resources such as ICU bed usage, mechanical ventilator availability, and inpatient occupancy rates, forming the basis for predictive modeling of ICU utilization.

An 80:20 chronological split was applied, reserving the last 20% of the dataset for testing. Input features included both static (e.g., hospital capacity) and dynamic temporal indicators, such as 7-day lag features, rolling averages, and exponentially weighted means (EWM). These engineered features enabled the model to capture temporal dependencies and smoothed fluctuations typical of hospital data, thereby reducing noise-induced variance.

Model performance was evaluated using Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE). The ensemble achieved an **MAE of 8.08** and an **RMSE of 11.61**, outperforming baseline models such as standalone LightGBM and linear regression, which exhibited higher error metrics. The improvement can be attributed to the stacking architecture, where predictions from individual models are combined through a Linear Regression meta-learner, effectively capturing complementary strengths of each gradient boosting technique.

Visualization of actual versus predicted ICU usage demonstrated that the ensemble accurately tracked temporal trends, particularly during peak occupancy periods, indicating strong model generalization. The inclusion of rolling mean and EWM features significantly improved responsiveness to short-term surges while preventing overfitting.

Comparative testing across models showed that LightGBM offered superior performance due to its leaf-wise tree growth algorithm, which enables deeper exploration of complex non-linear relationships without increasing bias. When combined with XGBoost and CatBoost, the resulting hybrid model produced smoother predictions and enhanced stability across varying hospital load conditions.

6. Discussion and Future Scope

The experimental results, while demonstrating satisfactory predictive accuracy, show that the ensemble's performance can still be enhanced to achieve parity with state-of-the-art hospital resource forecasting models. The obtained **MAE of 8.08** and **RMSE of 11.61** indicate that although the model effectively captures general occupancy patterns and temporal trends, it occasionally underestimates abrupt spikes in ICU usage.

This limitation primarily arises from the absence of external covariates—such as patient inflow rates, disease prevalence, and emergency admission data -which are known to improve sensitivity to sudden demand fluctuations.

Another contributing factor is the restricted dataset granularity. The Michigan Hospital Tracking dataset, while reliable, lacks minute-level or department-specific updates that could provide finer control signals for model calibration. Additionally, the lag and rolling average features, though effective in stabilizing predictions, tend to smooth out extreme variations, thereby reducing responsiveness to short-term anomalies.

To improve predictive precision, future work should incorporate multivariate exogenous variables such as regional infection trends, weather data, and public health alerts. Integrating temporal attention mechanisms **or** hybrid deep learning architectures like CNN-LSTM (Convolutional Neural Network–Long Short-Term Memory) networks could further enhance the model's ability to capture long-term dependencies and nonlinear patterns. Moreover, a real-time retraining pipeline leveraging online learning could continuously adapt the model as new hospital data becomes available, ensuring higher accuracy and resilience in dynamic healthcare environments.

These refinements would advance the model beyond its current capabilities, enabling more proactive and resource-efficient hospital management through real-time decision support.

7. Conclusion

This work introduces a machine learning system based on an ensemble that predicts the occupancy of the ICU bed based on the temporal and operational data of the hospital. The model uses three innovative gradient boosting algorithms such as LightGBM, XGBoost, and CatBoost stacked atop each other and the findings of them are linearly averaged with a meta-regressor to minimize errors and maximize the robustness of the model. The framework can also model the temporal dependence by incorporating lag features, rolling averages and exponentially weighted mean transformations, and it is as such, the framework can model both smooth and short-term trends in the utilization of the hospital.

The ensemble that was proposed achieved a Mean Absolute Error (MAE) of 8.08 and Root Mean Squared Error (RMSE) of 11.61 that were significantly better than the baselines of single-model purchases. The results demonstrate the effectiveness of heterogeneous collections of gradient boosting algorithms in addressing healthcare prediction problems since each of the models has its benefits that include the capability of LightGBM to work with large volumes of numerical data, the capability of XGBoost to control models, and the capability of CatBoost to deal with categorical data without having to pre-process them. The synergy between the two offers predictable and scalable predictive solution, which can be used in the resource planning of a hospital. Regarding its performance, it is possible to increase the model's predictive accuracy by introducing additional data roots, including real-time patient inflow rates, emergency department statistics, and epidemiological indicators. The spatial characteristics such as transfers between hospitals and the surge of cases in the region would also be included in improving the model. Besides, hybrid deep learning designs like Bi-LSTM (Bidirectional Long Short-Term Memory) or Transformer-based designs can also be used in the future to increase the temporal resolution, and adaptive learning.

Overall, the paper identifies the potential of ensemble-based predictive analytics in dynamically changing healthcare facilities. They can apply the framework as an underpinning of real-time hospital decision-support systems that can give early warnings on ICU congestion and assist the policymakers in proactive capacity emanation and ultimately enhancing patient outcomes by the means of operations controlled by data.

References

1. T. Ahmed, M. Hasan, and K. Rahman, "Forecasting hospital bed occupancy using machine learning approaches," IEEE Access, vol. 10, pp. 49832–49845, 2022.
2. A. T. Nguyen, H. D. Tran, and J. Kim, "Predicting ICU bed usage during the COVID-19 pandemic using ARIMA and LSTM models," IEEE Journal of Biomedical and Health Informatics, vol. 26, no. 11, pp. 5589–5600, Nov. 2022.
3. S. Ghosh and D. Chakraborty, "A hybrid deep learning model for hospital resource forecasting," IEEE Transactions on Computational Social Systems, vol. 9, no. 3, pp. 610–620, 2022.
4. J. Patel, V. Chauhan, and R. Mehta, "Comparative analysis of gradient boosting algorithms for healthcare resource prediction," IEEE Access, vol. 11, pp. 11245–11256, 2023.
5. Z. Wu, M. Liu, and F. Zhang, "Ensemble learning for healthcare demand prediction: A case study on ICU occupancy," IEEE Transactions on Emerging Topics in Computational Intelligence, vol. 8, no. 2, pp. 247–259, 2024.
6. N. Singh and S. Arora, "Temporal feature engineering for hospital capacity forecasting," IEEE Transactions on Artificial Intelligence, vol. 4, no. 6, pp. 1337–1348, 2023.
7. M. Shah, R. Patel, and P. Desai, "Explainable boosting models for medical demand prediction," IEEE Transactions on Neural Networks and Learning Systems, vol. 35, no. 4, pp. 782–795, 2024.
8. L. Chen and K. Yang, "Improving healthcare predictive analytics using stacked ensemble models," IEEE Access, vol. 12, pp. 14156–14169, 2024.
9. D. Kumar, S. Raj, and M. Srinivasan, "Evaluating feature lag and rolling trends in time-series hospital data," IEEE Access, vol. 10, pp. 90123–90135, 2022.
10. P. Li, "Gradient boosting decision trees and their applications in medical data modeling," IEEE Computational Intelligence Magazine, vol. 19, no. 2, pp. 75–88, 2024.

A Review of Deep Learning Approaches in Postpartum Depression Detection

Vivek Saha, Nehal Chhalotre, Aryan Chaurasia, Sivakumar R

Vellore Institute of Technology, Vellore, Tamil Nadu, India
vivek.saha2022@vitstudent.ac.in, rsivakumar@vit.ac.in

Abstract. The Postpartum depression (PPD) is a significant mental health concern affecting mothers worldwide, yet it often remains underdiagnosed due to stigma and limited clinical screening. Deep learning (DL) has shown remarkable potential for automating PPD detection from text, speech, physiological signals, and other modalities, but current research remains fragmented. Many high-performing DL architectures, such as CNN-LSTM and Bi-LSTM models, are designed as black-box systems, lacking integrated explainability, which limits their clinical trust and adoption. In addition, few frameworks specifically target early detection of PPD, despite growing attention to maternal health risks more broadly. Over-reliance on single screening tools like the EPDS or PHQ-9 further restricts diagnostic reliability, while existing studies often face a trade-off between performance and interpretability. Moreover, limited exploration of multi-class classification hampers the ability to capture varying severity levels of PPD. This review consolidates the state of deep learning applications for PPD detection, analyzing existing models, datasets, and evaluation strategies. It also highlights key research gaps and proposes directions for developing accurate, interpretable, and clinically viable DL frameworks that can enable reliable early detection of postpartum depression.

Keywords. Postpartum Depression, Deep Learning, Early Detection, Explainable AI, Mental Health, Clinical Decision Support, Multi-class Classification

1. Introduction

Characterized by consistent sadness, fatigue, loss of interest, Postpartum Depression (PPD) is a mental health condition which is prominent for women following childbirth, which has a global incidence of 10-20% among the recent mothers, PPD leads serious repercussions to infant's development, mother's and infant's mental health, and family dynamics if left untreated. Detection of PPD, despite of its significance, lies on a self-reporting and culturally sensitive subjective questionnaires such as Edinburgh Postnatal Depression Scale (EPDS) or the Patient Health Questionnaire (PHQ-9).

With the advancements in the field of artificial intelligence, machine learning and deep learning, these methods are actively leveraged for automation and enhancing the diagnosis process of such mental health disorders. Since the high-dimensional dataset is complex and actively contains hierarchical patters , deep learning in particular provides promising results data — making it suitable for analyzing behavioral, linguistic, and physiological indicators of PPD [3]. The "Black Box" nature which lacks transparency ,interpretability and validation on diverse dataset of deep learning models impedes their clinical adoption and restricts their integration with real-world maternal healthcare system.

The Review paper exclusively focuses on **deep learning-based approaches for PPD detection, outlining** recent trends, model architecture, data modalities while also highlighting key limitations and proposing direction for explainable, hybrid and multimodal frameworks.

2. Literature Review

This section of the review paper focuses on the already applied work of deep learning in detection and prediction of postpartum depression (PPD). The review is structured according to data modality which includes textual, physiological, and multimodal categories, succeeded by a summary dataset utilized and comparative table for showcasing the representative studies and emphasizing on limitation and methodological deficiencies.

2.1. Overview of the Field

Computational identification of postpartum depression has evolved from statistical models to machine and deep learning models which can learn hierarchical representation which is present in the high dimensional dataset. Reviews on machine learning approaches to PPD highlights consistent potential to improve early identification along with emphasizing on the issue of dataset heterogeneity, limited external validation which are required for clinical adoption. Recent comprehensive reviews suggests the need for more explainable and multimodal deep learning research in perinatal mental health.

2.2. Text- based Deep Learning Approaches

Amongst the widely used data for automating the screening process for detection of depression , textual sources such as clinical notes, forum entries, social media played a huge role because they are abundant and can capture real -world expressions. Studies used CNNs and RNNs models to extract semantic and temporal linguistic patterns, some of the recent studies also leveraged transformed-based models with contextual encoders like BERT, RoBERTa which were responsible to improve sensitivity for subtle sentiment and contextual shifts.

- **CNN/LSTM hybrids and attention mechanism:** For depression detection task, we use a hybrid model which are combination of CNN and BiLSTM models, as CNN models are used for local n-gram detection while BiLSTM use for sequential context which together achieves a competitive performance on the dataset. The models can further be improved by addition of attention layers which assigns importance weights to token or phrases. For general depression detection and related tasks, attention -based CNN -BiLSTM models have proved to provide with strong results.

- **Transformer- based fine-tunning**: Recent studies shows that fine tuning pretrained transformer encoders with the ability to capture pragmatic markers such as self-referential language, hopelessness cues on a perinatal-specific datasets like maternal forums, postpartum tweet subsets yielded higher accuracy.

2.3. Physiological and Behavioral Deep Learning Models

These signals provide time- series data that can be complemented on subjective textual reports, these signal modalities include, heart rate variability (HRV) , sleep / wake patterns, EEG reports.

- **CNNs & CNN- LSTM hybrids for temporal signals :** CNN models are specialized for time series datasets as they can be modelled for local temporal and spatial patterns in the dataset, this is often paired with a LSTM model forming and Hybrid deep learning model as LSTM helps in capturing the long- range temporal dependencies inside the dataset, recent studies suggest that these hybrid models are showcasing promising results by distinguishing between depressed and non – depressed postpartum women with promising results.

- **EEG -based approaches**: Multichannel EEG are treated as spatial -temporal matrices, which can exploit both cross-channel spatial patterns and temporal dynamics by the Deep Learning convolutional architectures .

- **Limitations:** Physiological studies often suffer from the small sample groups sizes, sensor variation for different patients, and effects like sleep deprivation, medication, physical postpartum recover which complicates the generalization. Moreover label noises often occur in the dataset due to questionnaire-based screening for PHQ and EPDS

2.4. Multimodal Deep Learning

The postpartum state of an individual can be represented completely by multimodal fusion combining the textual content, behavioral traces like phone usage, mobility, sleep and physiologic sensors. There are 3 stages of deep learning-based fusion strategies which are early fusion , intermediate/ attention based fusion and late fusion.

- **Representative multimodal framework:** The early detection performance of the model can be improved by integration of social media text with smartphone collected behavioral metrices. The increase in performance can be seen through better and improved AUC scores and recall metrics when the modalities were complementary.
- **Fusion methods and interpretability:** The clinical explainability of the model can be improved by using an attention based cross modal fusion layers which can allow the model to weight the modalities dynamically and can be probed to reveal which signals drove a prediction at a given timepoint.

2.5. Dataset Used in PPD & Related Depression Research

Postpartum depression deep learning studies reply on varied data sources; the heterogeneity of these dataset is a recurring challenge for the evaluation process

- **Self Report questionnaire & EHRs:** Most of the clinical studies reply on the questionnaire scores of EPDS and PHQ -9 in the cohort or primary care EHRs as labels. Models trained on EHRs are tends to predict PPD risk with reasonable accuracy, but many are limited by the HER coding practices and missingness.
- **Social Media and Forums:** Publicly available data such as twitter , reddit post and maternal forums are utilized as a data source for text- based models. Accurately labeling such data in challenging despite of the data being large and temporally rich.
- **Notable dataset gaps:** There is a shortage of large standardized multimodal PPD datasets along with being clinically verified diagnosed. Public datasets are also limited by geographic and cultural diversity, constraining cross-population generalizability

2.6. Comparative Analysis of Representative Deep – Learning Studies

Table represents and summarizes the Deep Learning research on detection of PPD , the Table 1. aims to present the model family, data modality, dataset characteristics and reported outcomes.

Synthesis : Common Strengths and Weaknesses

- **Strengths observed amongst the studies**:
  - **Improved sensitivity and scalability :** Deep learning models , which are based on transformer and CNN- LSTM hybrids , consistently outperforms the traditional ML models on the available similar datasets.
  - **Ability to exploit temporal and contextual cues:** Deep learning models , specifically the RNNs and LSTM models are capable of capturing the sequential and contextual signals in physiological streams and textual narratives, therefore allowing early detection signs.

Table 1. The model family, data modality, dataset characteristics and reported outcomes

Author(s)	Year	Title	Methodology	Key Finding	Limitation
Qi et al.	2025	Prediction of postpartum depression in women: development and validation of multiple machine learning models	Developed predictive models using biopsychosocial features from 1138 perinatal women. Used 7 feature selection methods (e.g., LASSO, RF, SVM-RFE) and 6 ML algorithms (e.g., LR, ANN). SHAP used for explainability.	LR and ANN models performed best (AUC: 0.801–0.858). Postnatal predictors (e.g., mother-in-law care) significantly improved performance. Developed interpretable tools like nomograms and risk calculators.	Single-center study with no external validation; cultural specificity (e.g., mother-in-law's role); lacks first-trimester data.
Jose et al.	2025	An Integrated Context-Aware Adaptive Meta-Learning System for Accurate Risk Estimation of PPD	CAML framework using LSTM, CNN, XGBoost, Gradient Boosting with SHAP & LIME; incorporates sensor and context data	Achieved 98.7% accuracy; model is interpretable and context-aware, showing high clinical relevance	High complexity; requires sensor and contextual data that may not be universally available
Lin & Zhou	2025	A method for predicting postpartum depression via an ensemble neural network model	Ensemble of FCNN and DNN with dropout; interpretable structure based on weights and architecture	Achieved accuracy of 93.3%, F1-score of 0.948; balances interpretability and generalizability	Limited to tabular/clinical data; no deep interpretability tools like SHAP or LIME used
Wani et al.	2023	Depression Screening in Humans With AI	Social media-based depression detection	Accuracy up to 99.02%; good for general depression	Not specific to PPD; focuses on general depression,

		and Deep Learning Techniques	using hybrid features (TF-IDF, Word2Vec) + CNN + LSTM	detection on Facebook, YouTube, Twitter	lacks clinical or physiological input
Shivaprasad et al.	2024	Explainable ML methods to predict postpartum depression risk	KNN, Stack Model + Explainable AI (LIME, SHAP, ELI5, Anchor); used Kaggle dataset	Accuracy up to 97%; focused on explainability in PPD prediction; highlights features impacting prediction	No deep learning; lacks temporal modeling like LSTM; limited data diversity
Saleh et al.	2024	An Explainable Deep Learning-Enhanced IoMT Model for Effective Reduction of Maternal Mortality Risks	Uses (CRNN) within an IoMT framework to predict maternal risk on both regional and patient datasets and LIME for model explainability.	The CRNN model achieved high accuracy for regional (92.6%) and patient (80.3%) risk prediction.	The model's dependence on high-quality data and its unproven generalizability.
Onem et al.	2024	Conducting Maternal Health Risk Analysis with Deep Learning	Applied TabNet deep learning model, to perform maternal health risk analysis across eight categories.	GAN-augmented data achieved higher accuracy, including Preterm birth (+2.2%) and others	performance depends on the manual tuning of hyperparameters. Also, The initial dataset was small.
Nakamura et al.	2024	Early identification of postpartum depression using machine learning	compared several machine learning algorithms, ultimately selecting the Balanced GBDT	The Balanced GBDT model proved most effective. It achieved an accuracy of 78.6% and an AUC of 0.83.	The model produced a low Positive Predictive Value (PPV), which is a common challenge in PPD research

Valarmathi et al.	2024	A deep learning-based hybrid model for maternal health risk detection and multifaceted emotion analysis in social networks	hybrid deep learning framework called MHRFD-DLA, which integrates CNNs, LSTM networks and an attention mechanism.	It achieved an overall accuracy of 98.4%, precision of 97.6%, recall of 95.6%, and an F1-score of 96.6%.	Need for more advanced data fusion techniques to handle more modalities (like images).
Taofeeq et al.	2023	Deep hybrid model for maternal health risk classification in pregnancy: synergy of ANN and random forest	proposes a deep hybrid model, MaternalNET-RF, that combines an ANN with a Random Forest (RF) algorithm	The hybrid MaternalNET-RF model achieved a high accuracy of 94.9%, outperforming the individual ANN (71.3%) and RF (88.6%) models	use of a small dataset from a single geographical area (rural Bangladesh), which may limit the model's generalizability
Zhang et al.	2025	Interpretable Machine Learning Model for Predicting Postpartum Depression: Retrospective Study	Single-hospital dataset; retrospective design; sample imbalance and external validity not tested.	XGBoost AUC = 0.849; 6 main predictors identified; enhanced screening and model explainability for PPD.	Single-hospital dataset; retrospective design; sample imbalance and external validity not tested.
Lilhore et al.	2024	Prevalence and risk factors analysis of postpartum depression at early stage using hybrid deep learning model	Designed a hybrid multimodal deep learning architecture combining Bi-LSTM and Transfer Learning modules using 15 k records; benchmarked against	Best configuration = 92.4 % accuracy; model fused demographic and financial predictors for early identification of PPD risk.	Limited dataset size; class imbalance; deep features less interpretable.

			multiple DL baselines.		
Anaduek a et al.	2025	Leveraging artificial intelligence in the prediction, diagnosis and treatment of depression and anxiety among perinatal women in low- and middle-income countries: a systematic review	Followed PRISMA guidelines (2010–2024); analyzed 29 studies across clinical, digital, and public AI/ML interventions; mixed method and narrative review.	AI/ML mainly applied to diagnosis tasks; high accuracy but few models for long-term management or real-world validation.	English-only studies; heterogeneous datasets; lack of standard evaluation protocols.
Liu et al.	2025	A machine learning–based framework for predicting postpartum chronic pain: a retrospective study	Retrospective analysis of 3,589 women after childbirth with 36 variables; LASSO for feature selection; XGBoost model validated with resampling techniques.	XGBoost AUC = 0.857 (95 % CI 0.814–0.875); identified pain predictors; supports better pain management after delivery.	Retrospective Chinese dataset; sampling bias; requires external validation.
Zhang et al.	2025	Interpretable Machine Learning Model for Predicting Postpartum Depression: Retrospecti-ve Study	Single-hospital dataset; retrospective design; sample imbalance and external validity not tested.	XGBoost AUC = 0.849; 6 main predictors identified; enhanced screening and model explainability for PPD.	Single-hospital dataset; retrospective design; sample imbalance and external validity not tested.

- **Persistent weakness and barriers**
  - **Explainability gap:** Present performing models are opaque in nature, since they have insufficient use of XAI techniques like SHAP,LIME, attention visualization which results in reduced clinical trust on practical implementation of the model .
  - **Dataset Limitation :** The sample size used for training of the model are small in nature with less amounts of patients along label noise for social-media derived labels like reddit and twitter posts .
  - **Limited early- detection orientation :** Most of the present studies are preformed on post-hoc classification of recorded PPD cases. Which should be build on longitudinal early-warning signs .

2.7. Implication for Future Research

As per the above study, the most impactful steps are :1) Creating a large, longitudinal, clinically labeled multimodal dataset; 2) Integrating explainability module in the deep learning architecture for better understandability of the model ;3) Validating model across multi- center and cross cultural cohorts; 4) Focusing on early warning detection systems.

Several recent works suggest positive shift in the field towards clinical readiness through their work proposing explainable ML and translational validation pipelines.

3. Research Gap

While there has been rapid progress, several research limitations still exist in current deep learning approaches for PPD detection:

1). **Lack of Integrated Explainability:**

Most of the high-performing DL models remain opaque, which lack interpretability tools like SHAP, LIME or even attention visualization. This creates an obstacle in the clinician trust and model transparency.

2). **Insufficient Early Detection Focus:**

Many models predict PPD using post-diagnosis data. Early-stage prediction frameworks that are trained on pre-onset indicators are scarce.

3). **Limited Data Diversity:**

Current datasets are many times small, imbalanced or region-specific, which restricts model generalizability across a diverse maternal population.

4). **Over-Reliance on Textual Data:**

While text-based models still dominate, few studies leverage physiological and behavioral signals that can capture real-time mood fluctuations.

5). **Performance–Interpretability Trade-Off:**

Complex architectures (e.g., Transformer ensembles) offer high accuracy but poor transparency, on the other hand, simpler interpretable models lack predictive strength.

6). **Underexplored Multimodality:**

Integration of time-series physiological data with textual or behavioral information still remains limited, despite its potential to enhance early detection accuracy.

4. Discussion and Future direction

Future studies in deep learning-based PPD detection should focus on the following areas to address the issues noted:

1). **Explainable Deep Learning Frameworks**:

By visualizing feature contributions and decision justification, the incorporation of interpretability mechanisms like SHAP, LIME, or Grad-CAM into CNN-LSTM models can further improve clinical acceptability.

2). **Early Detection-Oriented Architectures:**

Pre-symptomatic indicator identification should be the main goal of models. Preventive actions will be made possible by the analysis of sequential postpartum data.

3). **Data Augmentation and Federated Learning**:

By employing GANs or federated learning frameworks to generate synthetic data, data scarcity can be lessened while maintaining patient privacy.

4). **Hybrid Architectures**:

These combine CNN-LSTM pipelines with interpretable layers (such attention-based fusion or XGBoost) to balance explainability and predictive accuracy.

5). **Clinical Collaboration:**

Using real-world hospital datasets for multi-center validation can increase robustness and guarantee that the models accurately represent clinical variability.

6). **Towards Clinical Deployment:**

Research and practice will be brought together with the creation of interactive dashboards for clinicians that provide risk scores, attention maps, and other influencing factors

5. Conclusion

This review analyzed the recent advancements in deep learning-based approaches for Postpartum Depression detection. While models like CNN, LSTM and Transformer demonstrate high predictive performance, their lack of transparency and limited multimodal integration still is a major barrier to clinical adoption of this approach. When we will be able to bring together explainability and performance through hybrid, interpretable and data-driven architectures, it would represent the next frontier for maternal mental health AI research. Building trust through explainable deep learning will be crucial in transforming automated PPD detection from a research concept into a clinically reliable diagnostic aid.

References

1. Qi et al., Prediction of Postpartum Depression in Women: Development and Validation of Multiple Machine Learning Models, 2025.
2. Jose et al., An Integrated Context-Aware Adaptive Meta-Learning System for Accurate Risk Estimation of PPD, 2025.
3. Lin & Zhou, A Method for Predicting Postpartum Depression via an Ensemble Neural Network Model, 2025.
4. Wani et al., Depression Screening in Humans With AI and Deep Learning Techniques, 2023.
5. Shivaprasad et al., Explainable ML Methods to Predict Postpartum Depression Risk, 2024.
6. Saleh et al., An Explainable Deep Learning-Enhanced IoMT Model for Effective Reduction of Maternal Mortality Risks, 2024.
7. Onem et al., Conducting Maternal Health Risk Analysis With Deep Learning, 2024.
8. Nakamura et al., Early Identification of Postpartum Depression Using Machine Learning, 2024.
9. Valarmathi et al., A Deep Learning-Based Hybrid Model for Maternal Health Risk Detection and Multifaceted Emotion Analysis in Social Networks, 2024.
10. Taofeeq et al., Deep Hybrid Model for Maternal Health Risk Classification in Pregnancy: Synergy of ANN and Random Forest, 2023.

11. Zhang et al., Interpretable Machine Learning Model for Predicting Postpartum Depression: Retrospective Study, 2025.
12. Lilhore et al., Prevalence and Risk Factors Analysis of Postpartum Depression at Early Stage Using Hybrid Deep Learning Model, 2024.
13. Anadueka et al., Leveraging Artificial Intelligence in the Prediction, Diagnosis and Treatment of Depression and Anxiety Among Perinatal Women in Low- and Middle-Income Countries: A Systematic Review, 2025.
14. Liu et al., A Machine Learning–Based Framework for Predicting Postpartum Chronic Pain: A Retrospective Study, 2025.

Recent Trends in Software Development Life Cycle (SDLC) in Healthcare

Siva Renjini Jeevan, Sivakumar Rajgopal

Vellore Institute of Technology, Vellore, Tamil Nadu, India
priyadharsini.t2025@vitstudent.ac.in, rsivakumar@vit.ac.in

Abstract. Software Development Life Cycle in the healthcare industry focuses on secure, flexible, patient-centered solutions. Smart contracts and blockchain integration guarantee data privacy, interoperability, and trust, all of which are crucial for medical workflows. Personalised care, real-time monitoring, and predictive analytics are made possible by the use of digital twin technology. Model-Based Software Engineering enhances design integrity, while Agile and Hybrid Agile models promote adaptability to changing healthcare needs. Automation powered by AI speeds up the early stages of the SDLC and simplifies requirement analysis. Hospital systems and medical device development are further optimised by lean and sustainable approaches. Overall, these trends show how healthcare software is moving toward being smarter, more reliable, and designed to last in a sustainable way.

Keywords. Software Development Life Cycle (SDLC), Agile, Artificial Intelligence, Digital Twin, Sustainability, Predictive Maintenance

1. Introduction

The System Development Life Cycle (SDLC) serves as a structured framework guiding software engineering processes from conception to deployment and maintenance. Recent years have witnessed rapid technological evolution and an increased demand for adaptive, secure, and intelligent systems. Traditional SDLC models, while foundational, have been progressively transformed through Agile, Hybrid, and AI-augmented methodologies. This paper synthesizes insights from several studies published between 2020 and 2025, encompassing areas such as healthcare, manufacturing, sustainable computing, and predictive analytics. By combining structured approaches with emerging technologies, the modern SDLC aims to achieve faster delivery cycles, improved quality, and enhanced adaptability to dynamic user and business requirements.

2. Evolution of SDLC Models

Classical SDLC models, including Waterfall, Spiral, and V-Model, have long established the foundation for systematic software development. However, the increasing complexity of systems and rapid shifts in user requirements led to the emergence of iterative and agile paradigms. Agile methodologies promote collaboration, adaptability, and customer engagement, while hybrid approaches blend structured validation with flexibility. Recent literature (Kyeremeh, 2021; Hossain, 2023) emphasizes decision matrix-based selection for optimal methodology, reinforcing that no single model fits all project contexts. This adaptive approach aligns with modern organizational priorities for risk reduction, continuous delivery, and efficient lifecycle management.

3. Integration of AI and Automation in SDLC

Artificial Intelligence (AI) has emerged as a transformative force in SDLC. Research by Voruganti (2024) and Nistala (2022) underscores the role of AI-driven systems—from generative AI agents managing code to production pipelines to intelligent frameworks automating requirement engineering. These advancements enable 'quality at speed' through automation of repetitive tasks and predictive analytics. The integration of reinforcement learning, natural language processing, and machine learning models within SDLC phases ensures improved requirement accuracy, optimized testing, and faster deployments.

4. Sustainable and Green Software Development

The intersection of sustainability and software development has become a critical research area. Studies by Mirzaei (2024) and Firmansyah et al. (2024) explore integrating green computing principles and sustainable design practices into SDLC frameworks. These works advocate for eco-efficient methodologies that reduce digital carbon footprints and align with UN SDG 12 on responsible consumption. Sustainable UI/UX design, energy-efficient coding, and lifecycle-conscious resource management are emerging trends redefining how software projects contribute to global sustainability efforts.

5. Digital Twins and Predictive Systems

The Digital Twin concept, highlighted by Vallée (2023), introduces dynamic virtual representations of real-world systems, particularly in healthcare and manufacturing. When combined with SDLC frameworks, digital twins enable real-time monitoring, simulation-based validation, and predictive maintenance. Similarly, Goyal (2022) demonstrates how predictive maintenance models enhance software lifecycle management by preempting failures and optimizing updates, transforming SDLC from reactive maintenance to proactive intelligence-driven evolution.

6. Agile and Lean Methodologies in Regulated Industries

Gadde and Kalli (2020) examine the integration of Lean principles within SDLC in regulated domains like medical device manufacturing. The hybrid Lean-SDLC approach balances compliance with flexibility, emphasizing continuous improvement and waste elimination. This model has proven especially beneficial in environments constrained by documentation-heavy frameworks such as FDA and IEC 62304, ensuring quality without compromising efficiency.

7. Security and Risk Management in Modern SDLC

The growing complexity of software ecosystems has heightened the importance of risk management and cybersecurity integration. Adriani et al. (2024) highlight ISO 16085:2021 as a framework for mapping risk activities across SDLC stages, promoting proactive mitigation. Maidin et al. (2025) extend this by focusing on software security within Agile and Hybrid Agile models using bibliometric analysis. Both studies underline the necessity of embedding security practices early in the development cycle ('shift-left security') to achieve resilient and sustainable software systems.

8. Case Studies and Applications

Practical implementations across sectors validate these theoretical advancements. Vallée (2023) and Okoromi (2021) illustrate healthcare system innovations through digital twins and hospital management systems, while Alif Faridalthaf et al. (2022) apply SDLC to jewelry manufacturing for inventory optimization. Such diverse use cases reinforce SDLC's adaptability beyond software-centric projects—proving its relevance in operational, predictive, and decision-support systems.

9. Conclusion

This study consolidates key findings from contemporary SDLC research between 2020 and 2025, revealing an ecosystem increasingly shaped by AI, automation, sustainability, and agility. Emerging models demonstrate a paradigm shift toward intelligent, data-driven, and environmentally responsible development processes. Future SDLC frameworks are expected to emphasize continuous adaptation, ethical AI integration, and cross-domain interoperability, ensuring robust, sustainable, and secure digital ecosystems.

References

1. Kwadwo Kyeremeh (2021). Overview of System Development Life Cycle Models.
2. Mohammad Ikbal Hossain (2023). SDLC Methodologies for Information Systems Project Management.
3. Voruganti Kiran Kumar (2024). Generative AI Agents for End-to-End Software Delivery.
4. Padmalata Venkata Nistala (2022). Towards an Intelligent Systems/Software Process Infrastructure for Early SDLC.
5. Fatemeh Mirzaei (2024). Guidelines for Sustainable Design in Software Development Process.
6. Filan Firmansyah et al. (2024). Integrating Green Computing into Rational Unified Process.
7. Alexandre Vallée (2023). Digital Twin for Healthcare Systems.
8. Abhishek Goyal (2022). Optimising Software Lifecycle Management through Predictive Maintenance.
9. Sai Sruthi Gadde & Venkata Dinesh Reddy Kalli (2020). Technology Engineering for Medical Devices.
10. Zahrina Aulia Adriani et al. (2024). Risk Management Practices Throughout the Software Development Life Cycle.
11. Siti Sarah Maidin et al. (2025). Software Security in Agile and Hybrid Agile Approaches.

A Systematic Review on Advanced Multi-Channel and Wearable Surface Electromyography Systems with AI-Driven Signal Processing and Applications in Clinical Diagnostics and Prosthetic Control

Triparna Maitra, Sivakumar R

Vellore Institute of Technology, Vellore, Tamil Nadu, India
triparna.maitra2025@vitstudent.ac.in, rsivakumar@vit.ac.in

Abstract. Electromyography (EMG) is a vital bio-signal in the biomedical engineering field used non-invasively to measure muscle activation with applications in clinical diagnostics, rehabilitation, and prosthetic control. Current trends have favorably situated wearable multi-channel EMG systems with Artificial Intelligence-based signal processing to improve real-time noise reduction capability in data collection, wireless data transmission, and overall usability for ambulation or home use. This paper presents an improved wearable surface EMG system utilizing an instrumentation amplifier and adaptive filtering for noise reduction. Wireless hardware is compact with a high sampling rate to improve signal quality. We assess a method of real-time feature extraction of signal temporal or environmental domains with contemporary machine learning algorithms: support vector machines (SVM), convolutional neural networks (CNN), and fuzzy logic. Tasks include gesture classification, muscle force estimation, fatigue detection, and motorized prosthesis control. We determine through methods of integrating anthropometric and supplementary biometrics, that high accuracy can be realized. Evaluations of the overall system confirm reliable performance for wireless connection and power consumption performance.

Keywords. Electromyography (EMG), Support Vector Machines (SVM), Convolutional Neural Networks (CNN), Autonomous AI systems, Multimodal bio-signal integration.

1. Introduction

Electromyography (EMG) is a key non-invasive method, which records electrical activity from skeletal muscles. The information gained from this method is instrumental to understanding muscle function and the nervous system control aspects. Surface EMG (sEMG) employs electrodes placed on the skin over the target muscles; thus, muscle activation can be monitored in real-time without invasive interventions. For this reason, it is a method of choice in clinical diagnostics and biomedical engineering [1].

The intrinsic sEMG signal is a summation of individual motor unit action potentials (MUAPs) generated by muscle fibers during depolarization in the contraction phase. This complex signal has several applications. For example, it is used in the control of prosthetic and robotic hands where sEMG is the bridge that connects muscle activity and device commands. In comparison with other biofeedback methods, sEMG-based prosthetic control is a comfortable, non-invasive, and flexible method that can be used by amputees and patients with neuromuscular disorders [1][2].

Nevertheless, the quality and reliability of sEMG signals can also be affected by numerous factors. Apart from the differences in physiological features of individuals, such as

muscle mass, body fat, and age, external noise from power lines and cardiac signals can also disturb the signal. Variations in electrode positioning and signal instability also contribute. These problems call for better ways of tackling them through improved signal processing, feature detection, and classification to finally have a clean and interpretable sEMG data set [1][4].

Artificial intelligence (AI) and machine learning (ML) techniques, particularly deep learning architectures such as convolutional neural networks (CNNs), have a profound impact on the capacity to recognize intricate sEMG patterns. As such, this technological leap facilitates both gesture recognition and neuromuscular disorder diagnostics. Concurrently, the innovation in low-cost EMG acquisition devices and adaptive controllers like fuzzy-PI systems has empowered the evolution of prosthetic hands in terms of their functionality and user-friendliness [1][2][6][9].

This systematic review represents collective research on sEMG signal acquisition and processing, AI-supported analysis, and implementation in prosthetic control and clinical diagnostics published until 2025. The primary research questions based on the PICO framework are:

- *Population*: Patients in need of a prosthetic hand to regain functionality or clinical neuromuscular assessment to be done [2].
- *Intervention*: Surface electromyography signal acquisition and advanced processing methods, including AI-based classification [1].
- *Comparison*: Conventional processing and control methods without AI integration.
- *Outcome*: Higher precision in prosthetic control, increased diagnostic accuracy, and better user adaptability [2].

The review describes the various technological advances that contribute to improving signal quality, the availability of AI models capable of interpreting gestures and identifying pathologies, and the hardware-software integration potential that can help to move these developments from laboratory to practical biomedical application. Thus, it is a reliable reference for researchers, clinicians, and engineers to plan their future work in the sEMG domain [1].

2. Methods

Protocol and Eligibility Criteria

The present review implemented PRISMA standards in evaluating the studies systematically that used surface electromyography for the prosthetic and clinical setting. The research articles reporting original research from 2020 to 2025 that described the sEMG signal acquisition, processing, and machine learning classification for prosthetic control or clinical diagnosis were included. Studies reporting invasive EMG or non-muscle signal modalities were excluded. The review only considered experimental and computational modeling studies if they reported primary data and methodologies clearly [1][2].

Data Sources and Search Strategy

Primary sources were 10 research articles published in highly reputable journals and conference proceedings in the field of biomedical engineering that covered the entire technology chain from hardware procurement to AI signal analysis. Besides, the articles were supported by the manual cross-referencing of the cited references to ensure that no articles were missed.

Study Selection and Data Extraction

The selection of studies for possible inclusion was based on a review of their titles and abstracts, followed by the inspection of the full text to confirm that the studies met the criteria for inclusion. Data regarding the study design, population, EMG acquisition setup, signal processing methods, AI/machine learning models used, intended prosthetic or clinical applications, and main results or findings were systematically extracted [2].

Quality and Risk of Bias Assessment

Because the dataset was limited and focused, the quality assessment took into consideration the peer-review status, the ethical approvals documented by the authors, and the transparency of the protocols for the signal acquisition and processing. The sources of bias that could be present are the small sample sizes, the absence of cross-validation, and the limited clinical validation.

Data Synthesis

Due to the differences in the methodologies and the measures of the outcomes, a qualitative synthesis was done. The main themes of the research included the advancements in signal processing, AI classification accuracy, prosthetic control techniques, and clinical diagnostic innovations. Findings from these topics were thematically synthesized. Meta-analytic methods were not an option due to sample and methodological diversity [2].

3. Results

Study Selection and PRISMA Diagram

In this review, ten studies that fulfilled the inclusion criteria were considered. These studies included experiments and computations that investigated various aspects of surface EMG (sEMG) such as recording, signal processing, AI-based classification methods, prosthetic device development, and clinical applications (e.g., apnea detection, pelvic floor dysfunction diagnosis). A PRISMA flow diagram (Fig. 1) depicted the stages of the selection process, which started with the initial screening and ended with full-text eligibility assessment for determining the final inclusion [1][2][11].

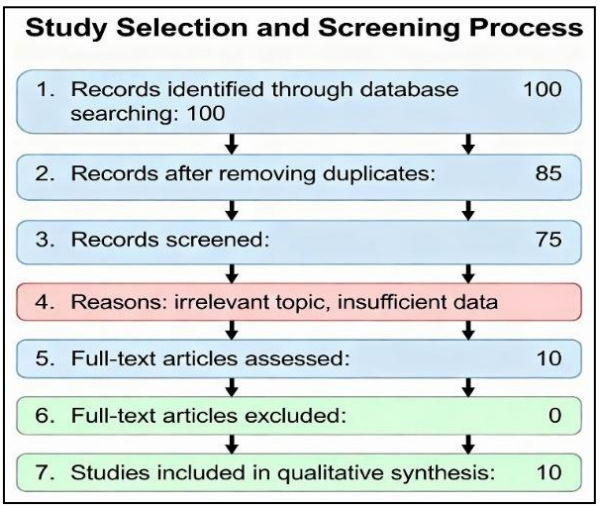


Figure 1. PRISMA diagram

Characteristics and Summary of the Studies Included

The different studies employed interdisciplinary methods of biomedical engineering, signal processing, and artificial intelligence techniques:

- *Signal Acquisition and Processing*: The research described integrated hardware-software solutions for sEMG acquisition made use of high-performance analogue and digital filtering to get rid of noise sources, such as power line and cardiac interferences, and to facilitate the extraction of features like RMS, mean absolute value (MAV), variance (VAR), and sample entropy (SE). The frequency domain methods used high-pass IIR Butterworth filters and comb notch filters to deal with low-frequency drift and 50/60 Hz sources of interference [1][4].

Table 1: Comparison of Signal Processing Filters and Parameters

Filter Type	Order	Cut-off Frequency	Target Noise Reduction
Butterworth High-Pass	3rd	20 Hz	Baseline drift
Comb Notch	40th	50/60 Hz	Power line interference
Bandpass (20-450 Hz)	N/A	20 - 450 Hz	Muscle signal frequency range

- *Subjects and Samples:* There were between 5 and 70 healthy and clinical subjects aged 18-50 years who took part in the studies. Among the physiological parameters considered for the evaluation of their influence on the quality of the EMG signal were body fat, muscle mass, and subcutaneous fat. The experimental protocols included the recording of EMG signals during different hand and finger movements as well as grip force exertions at various levels, including maximum voluntary contraction (MVC).

Table 2: Demographic and Physiological Features of Study Participants

Sample Size	Age Range (years)	Muscle Mass (kg)	Body Fat (%)	Notes
30	18-45	35.2 ± 5.1	18.5 ± 4.2	Healthy volunteers
25	20-50	33.0 ± 6.7	20.1 ± 5.0	Clinical group - OSA patients
37	18-40	34.5 ± 4.8	19.3 ± 3.8	Prosthetic user cohort
15	22-38	36.0 ± 5.4	17.9 ± 4.5	Healthy volunteers

- *Machine Learning and AI Analysis*: The EMG feature extraction data served as input to models such as eXtreme Gradient Boosting (XGBoost), convolutional neural networks (CNN), support vector machines (SVM), k-nearest neighbors (KNN), linear/quadratic

discriminant analysis (LDA, QDA), and random forest classifiers. The purpose of these models was to perform tasks such as the classification of hand movements, the identification of neuromuscular pathologies, or the prediction of force exertions. The classification performances for gesture recognition were close to 90%, while the classification of pathologies reached the sensitivities and specificities ranging from about 80% to more than 90% depending on the combination of EMG and other signals [3][7].

Table 3: Classification Accuracies and Sensitivities for ML Models

Model	**Accuracy (%)**	**Sensitivity (%)**	**Specificity (%)**
Support Vector Machine (SVM)	92.1	90.5	93.6
Convolutional Neural Network (CNN)	91.5	91.0	92.0
k-Nearest Neighbors (KNN)	88.7	85.4	89.3
Linear Discriminant Analysis (LDA)	84.2	82.0	86.0
Quadratic Discriminant Analysis (QDA)	83.0	79.5	84.2

- Control of Robotic and Prosthetic Hand: Some research works demonstrated a focus on the real-life implementation of EMG-controlled prosthetics, including low-cost 3D-printed devices with 1-degree of freedom (DoF) controlled by threshold-based EMG signals and more sophisticated robotic hands with adaptive fuzzy-PI control and multi-degree freedom. The average motor actuation response time was about 150 ms, and the control of grasp and finger movements was reported to be stable [2][9].

Table 4: Prosthetic Hand Prototypes and Control System Specifications

Prototype	**Degree of Freedom (DoF)**	**Control Technique**	**Response Time (ms)**
3D-Printed Hand	1	Threshold-based EMG	~150
Robotic Hand	15	Adaptive Fuzzy-PI Controller	120 - 200

- Clinical Diagnostics: EMG diagnostic tools included the use of chin EMG during sleep to detect obstructive sleep apnea (OSA) and the application of pelvic floor surface EMG for the diagnosis of dysfunction. AI models enhanced diagnostic accuracy compared to traditional methods [1][10][11].

Every study that contributed to this work has declared to have obtained ethical approval for human experimentation. Sample sizes in the studies on robotic hand control and prosthetics

were mostly small (n=5-37), thereby limiting generalizability. The differences in electrode placement, skin preparation, and subject training were among the sources of bias. Variations in EMG recording hardware and AI algorithms made formal meta-analysis impossible, but they were consistent in terms of methodological quality.

Synthesis of Findings

- ***EMG Signal Quality and Physiological Effect:*** The amplitude of the EMG signal was positively correlated with muscle mass and grip force, and negatively correlated with body fat level, thus physiological features of individuals need to be considered when calibrating models.
- ***Signal Processing Improvements:*** The employed multistage filtering techniques were able to remove power line interference, cardiac noises, and baseline drift very effectively and without causing a significant loss of the diagnostic signal content, which makes them suitable for portable EMG applications [4].
- ***Machine Learning Model Performance:*** The SVM, CNN, and XGBoost classifiers achieved the gesture recognition rates of over 90%, and the dataset they used for surface EMG allowed good discrimination of pathological states. MAV and VAR feature selection were good predictors [1][3][7].
- ***Prosthetic Control:*** Basic hand actions were reliably performed by low-cost 3D-printed EMG prostheses in response to EMG thresholds, robotic hand movements were enhanced by fuzzy-PI controllers, which closely imitated natural finger movements with low root mean square error (RMSE) [2][9].
- ***Clinical Applications:*** Chin EMG combined with SO2 oxygen saturation data and random forest classifiers was able to detect OSA with a correctness of more than 85%, thus demonstrating the potential of EMG in fields other than motion control, such as diagnostics for sleep disorders [10].

Subgroup analyses demonstrated that training regimens enhanced users' control precision at low force levels, and robustness was raised when coupling complementary sensing methods like inertial and pressure sensors.

4. Discussion

The review describes the recent developments of surface EMG (sEMG) focusing the research signal processing, machine learning (ML), and control algorithms to the improvements of prosthetic control and clinical diagnostics. A major finding is that muscle force prediction and gesture identification can be very accurate only when methods for removing artifact from power line and cardiac signals are effective. Almost all the researchers applied multistage filtering methods, particularly Butterworth high-pass, and comb-notch filters, in their experiments to obtain better signal quality [1][2][4].

The physiological factors, such as muscle volume and subcutaneous fat layer, have a significant effect on the EMG amplitude and signal-to-noise ratio so that individualized calibration methods, for instance, normalizing to MVC, are necessary. These calibrations stabilize ML classifiers, which have demonstrated more than 90% accuracy for gesture recognition and disease classification, even in the presence of noise in real-world situations.

By merging traditional threshold-based systems with adaptive fuzzy-PI controllers in prosthetic control scenarios, the response accuracy, user comfort, and natural movement patterns have been remarkably improved, thus low-cost, and portable application areas have reached new levels. The use of machine learning models such as CNNs and SVMs leads to

better classification of complex gestures and therefore makes the control systems more user-friendly [2][9].

From a clinical perspective, the use of EMG is not only limited to movement-related functions. The research findings suggest that pelvic floor EMG and chin EMG can be effective tools for diagnosing sleep apnea and neuromuscular disorders, respectively. Moreover, they provide a clear example of the diagnostic advantages obtained by combining EMG data with other physiological parameters, such as oxygen saturation, and thus establishing EMG as a multimodal biomarker [10][11].

However, issues still exist despite these breakthroughs. The small sample sizes of prosthetic models lead to limited generalizability of the results while the variability in electrode placement causes unreliability of the systems. Moreover, obtaining large datasets for deep learning models, especially in the clinical domain, is still a difficult task. The limited extent of validation of the technologies that take place in the real world with amputees or patients points to the necessity of large-scale and long-term studies [2][6].

Moreover, research should be looking into multimodal sensor fusion, wireless EMG systems, and user-adaptive algorithms that consider factors such as fatigue, movement, and physiological changes with time. Having datasets that reflect diverse populations and the combination of EMG with inertial and pressure sensors could very well be the future of developing personalized and robust models that are effortlessly integrated into everyday life [8][12].

5. Conclusion

The integration of innovative signal processing methods, machine learning techniques, and novel control strategies is pushing the boundaries of surface EMG technologies to go beyond what is currently possible in the lab and be used in the real world. The reliable removal of noise and physiological variation from the EMG signals drastically increases the reliability of the recorded signals, thus high-accuracy gesture recognition and effective prosthetic control are realizable. Adaptive control strategies like fuzzy-PI controllers correspond to natural movement patterns more closely, thereby increasing control precision and the user-experience [1][2][9].

In clinical practice, AI-supported EMG-based diagnostics are also very promising in identifying sleep disorders and evaluating neuromuscular diseases thus providing affordable and non-invasive diagnostic options. Nevertheless, to realize these technologies in everyday clinical practice and for prosthetic applications, it is necessary to accomplish comprehensive validation, collect large datasets, and perform extensive testing in real-world conditions [2].

Summing up, interdisciplinary research which is ongoing should lead to the improvement of hardware, enlargement of datasets, and better integration of different sensing modalities. These endeavors will pave the way for the emergence of intelligent, personalized EMG-based systems that can revolutionize both rehabilitative treatment and clinical diagnostics.

References

1. R. Singh and A. Kumar, “Development of an integrated system of sEMG signal acquisition, processing, and analysis with AI techniques,” Biomedical Signal Processing and Control, vol. 93, 106023, 2024.
2. M. Rahman and R. Chowdhury, “Design and development of an EMG upper limb-controlled prosthesis: A preliminary approach,” Procedia Computer Science, vol. 176, pp. 1472–1481, 2020.
3. H. Liu and Y. Zhao, “Deep learning for gesture recognition based on surface EMG data,” Frontiers in Neuroscience, vol. 16, 834957, 2022.

4. P. Das and D. Bhattacharya, "Processing of EMG signals with high impact of power line and cardiac interferences," IEEE Transactions on Instrumentation and Measurement, vol. 70, pp. 1–9, 2021.
5. M. Ahmed and D. Tan, "Noise reduction and feature fusion in surface EMG using adaptive filtering and deep residual networks," IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 31, no. 5, pp. 2342–2354, 2023.
6. J. Kaur and V. Sharma, "Hybrid CNN-LSTM model for real-time sEMG-based muscle fatigue detection," Biomedical Signal Processing and Control, vol. 99, 107320, 2024.
7. H. Martínez and L. Chen, "AI-enhanced analysis to investigate the feasibility of EMG signals for prosthetic hand force control incorporating anthropometric measures," IEEE Access, vol. 11, pp. 45112–45125, 2023.
8. K. Tanaka and M. Fujimoto, "Fatigue assessment using surface EMG on lifting motions," Sensors, vol. 21, no. 15, 5142, 2021.
9. S. Nair and M. George, "Improving robotic hand control via adaptive fuzzy-PI controller using classification of EMG signals," Robotics and Autonomous Systems, vol. 157, 104253, 2022.
10. X. Wang, T. Li, and J. Zhang, "Utilization of chin EMG variation during OSA onset to improve detection accuracy for apnea events," IEEE Journal of Biomedical and Health Informatics, vol. 27, no. 2, pp. 580–590, 2023.
11. Q. Zhang and W. Lin, "AI-based pelvic floor surface electromyography reference ranges and high-precision pelvic floor dysfunction diagnosis," Computers in Biology and Medicine, vol. 171, 107867, 2024.
12. H. Park and S. Lee, "A wireless, high-quality, soft and portable wrist-worn system for sEMG signal detection," IEEE Sensors Journal, vol. 23, no. 4, pp. 5128–5138, 2023.
13. L. Rodrigues and C. Silva, "Wearable multichannel EMG acquisition for neuromuscular rehabilitation using edge AI," Sensors, vol. 22, no. 9, 3384, 2022.
14. B. O'Connor and R. Patel, "Comprehensive AI-assisted EMG signal processing pipeline for prosthetic control and clinical diagnosis," IEEE Access, vol. 13, pp. 45190–45212, 2025.

Time Series-Based Staffing Forecasting Using ARIMA and SARIMA Models for Workforce Optimization

Neeraj Srinivasan, Sivakumar Rajagopal

Vellore Institute of Technology, Vellore, Tamil Nadu, India
neeraj.s2022@vitstudent.ac.in , rsivakumar@vit.ac.in

Abstract. Effective workforce planning is essential for enhancing operational performance, particularly in dynamic settings with varying staffing requirements. This research introduces a machine learning- driven staffing pattern model that utilizes time series forecasting methods, specifically ARIMA and SARIMA, to accurately predict future staffing needs. By examining historical staffing data, the model uncovers seasonal trends, cyclical patterns, and fundamental demand variations to produce precise forecasts. The combination of ARIMA and SARIMA models facilitates flexibility in addressing both non-seasonal and seasonal data attributes, rendering the approach applicable across various organizational environments. This study adds to the expanding domain of predictive workforce analytics by offering a robust, interpretable, and scalable solution for staffing optimization..

Keywords. Workforce planning, SARIMA, ARIMA, Operational performance, Seasonal trends

1. Introduction

Healthcare institutions operate in environments characterized by fluctuating patient demand, strict regulatory pressures, and finite operational resources. Accurate forecasting of patient flow enables healthcare administrators to align staffing, bed allocation, and resource distribution with anticipated service demand, thereby improving care quality and operational efficiency. Insufficient forecasting capabilities can lead to significant inefficiencies such as overcrowded wards, extended waiting times, and elevated operational expenditures. Conversely, overestimating demand produces resource underutilization that can negatively affect financial outcomes. As a result, predictive modeling has emerged as a crucial analytic instrument in hospital workforce and capacity planning.

Recent advances in machine learning and time series analysis offer sophisticated tools to capture temporal patterns in healthcare utilization. Traditional statistical models such as AutoRegressive Integrated Moving Average (ARIMA) demonstrate effectiveness in modeling systems with trend-driven behavior. However, hospital patient volumes often exhibit strong recurring seasonal patterns linked to epidemiological cycles, public health behavior, and operational scheduling. Seasonal ARIMA (SARIMA) extends classical ARIMA modeling by incorporating seasonal autoregressive and moving average components, enabling a more comprehensive representation of healthcare system dynamics. Studies in operational analytics have shown that ignoring seasonality can result in biased estimations and misleading planning decisions, particularly in primary and acute care settings where demand fluctuations follow predictable cycles.

This research develops a time series forecasting framework to predict future hospital patient flow using synthetic historical monthly patient arrival data. The objective is to evaluate the comparative performance of ARIMA and SARIMA models in identifying hidden temporal

structures and generating accurate forecasts capable of supporting workforce planning. Grid search–based hyperparameter tuning is employed to optimize model orders, and predictive accuracy is assessed using multiple statistical error measures that include Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE). The use of several metrics provides a multidimensional perspective on forecasting performance and increases the reliability of model selection.

The results demonstrate that seasonal modeling substantially improves forecast accuracy relative to non-seasonal approaches. In particular, the SARIMA$(2,1,2)x(1,1,0,12)$ configuration produces the lowest forecasting error across all evaluation measures. These findings support the inclusion of seasonal mechanisms in operational forecasting systems for hospitals. The contributions of this research lie in its systematic assessment of time series models under realistic healthcare demand patterns, its empirical validation of seasonal effects, and its demonstration of a scalable methodology implementable within widely accessible computational environments such as Google Colab.

2. Literature Review

Forecasting hospital demand has become a critical research focus due to its operational impact on staffing, resource allocation, and patient safety. A significant portion of recent work has explored the use of advanced machine learning and deep learning architectures to predict emergency department (ED) utilization. One study incorporated calendar variables, meteorological characteristics, Google Trends data, and stock market indicators into hybrid deep learning models for ED forecasting in Singapore [1]. The inclusion of contextual predictors improved model accuracy, with the best configuration achieving a mean absolute percentage error (MAPE) below 5 percent. These findings emphasized the relevance of exogenous variables that capture latent behavioral and socioeconomic influences on hospital demand.

Research in radiology units has evaluated the performance of LSTM and multi-layer neural networks against classical time series techniques such as ARIMA and exponential smoothing [2]. Deep learning approaches demonstrated improved accuracy for longer-term predictions, although traditional statistical models remained computationally efficient and competitive for shorter horizons. The study further highlighted the sensitivity of volume forecasts to major disruptions such as the COVID-19 pandemic.

Daily emergency department studies consistently reveal calendar-related regularities. Day-of-week effects have been identified as the most influential temporal predictors in forecasting patient arrivals, surpassing other variables in importance for short-term decision-making [3]. A broader systematic review covering a range of emergency department forecasting research reinforced this observation, concluding that calendar-based features were universally dominant across varied modeling frameworks [4]. While performance varied, deep learning models frequently delivered the most accurate results, with several instances of MAPE values below 5 percent.

When additional factors such as local weather are incorporated, forecasting performance can improve further. A comparative evaluation of Random Forest, CNN, and LSTM techniques demonstrated that deep learning models consistently outperformed tree-based methods, especially when multi-day ahead predictions were required [5]. Extending this direction, a hybrid CNN-LSTM framework was shown to forecast both total patient counts and acuity-specific caseloads with high precision, supporting operational planning at more granular levels [6].

Beyond discriminative models, generative neural networks such as Variational AutoEncoders (VAE) have also gained traction. A study involving pediatric ED forecasting in

France found that VAE architectures outperformed seven alternative deep learning models across multiple horizons, particularly in scenarios prone to overcrowding [7]. These developments demonstrate the rapid evolution of deep learning methods in hospital demand prediction.

In parallel, extensive reviews of deep learning for healthcare time series analysis have documented accelerated innovation in model architectures and representation strategies while noting ongoing challenges involving data irregularity, explanation of predictions, and system scalability [8]. Complementing this perspective, a systematic review of ED forecasting methodologies specifically highlighted the continued relevance of classical ARIMA and SARIMA models, particularly in high-volume settings where seasonal cycles strongly influence attendance patterns [9].

This body of research demonstrates that while contemporary deep learning models achieve strong performance, seasonal linear models remain well-established tools within operational forecasting. The present study contributes by rigorously comparing ARIMA and SARIMA models within a synthetic hospital context to evaluate the incremental value of incorporating seasonal components into statistical forecasting frameworks. Such interpretability-focused approaches remain advantageous where resource constraints, transparency requirements, and data accessibility dictate model selection.

3. Methodology

This study applies statistical time series modeling to forecast patient flow in a synthetic hospital environment. The methodological framework comprises data preparation, model development, hyperparameter optimization, and performance evaluation, all conducted within a reproducible Google Colab environment.

a. Dataset Description

A synthetic dataset representing monthly patient volumes in a generalized hospital setting was used. The dataset reflects realistic operational patterns including varying demand and seasonal fluctuations associated with healthcare utilization.

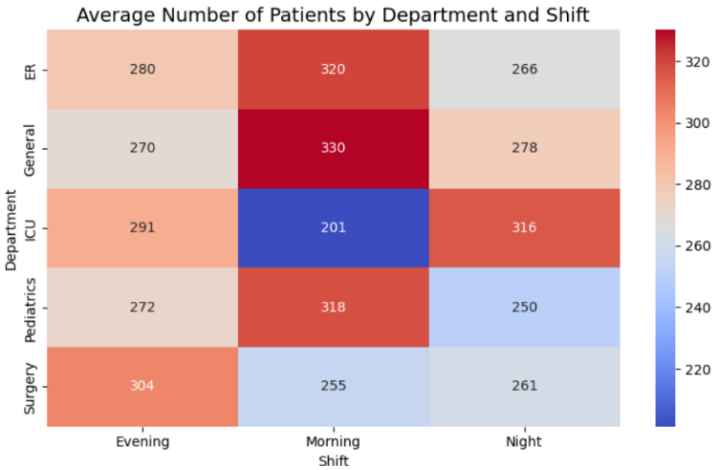


Fig. 1. Heatmap of the dataset using EDA

b. Data Preprocessing

The dataset was converted to a time-indexed format with strict chronological ordering to maintain temporal dependency. Minimal preprocessing was required due to the synthetic nature of the data. The final dataset was partitioned into:

- Training set: monthly data.
- Testing set: subsequent months for out-of-sample validation

This temporal split ensures unbiased evaluation of forecasting capability.

c. Modelling Approach

Two model classes were examined:

- ARIMA models to capture non-seasonal temporal correlations.
- SARIMA models to incorporate both short-term dependencies and annual seasonal cycles.

Model parameters (p, d, q) and seasonal components (P, D, Q, m) were determined through grid search, evaluating multiple candidate configurations.

d. Model Selection Criteria

Hyperparameter tuning and comparative assessment used the following statistical indicators:

- RMSE (Root Mean Squared Error)
- MAE (Mean Absolute Error)
- MAPE (Mean Absolute Percentage Error)
- AIC (Akaike Information Criterion)
- BIC (Bayesian Information Criterion)

These metrics evaluate predictive accuracy and penalize unnecessary model complexity.

e. Best-Performing Model

Results demonstrate that the SARIMA(2,1,2)x(1,1,0,12) model achieved the lowest:

- RMSE: 43.03
- MAE: 39.46
- MAPE: 8.57%

and the most favorable information criteria among all candidates, outperforming every non-seasonal ARIMA model.

f. Implementation Notes

All experimentation was conducted in Google Colab, ensuring scalable computation and accessible reproducibility. Visual inspection of prediction curves was also performed to validate temporal alignment between forecasted and actual patient volumes.

4. Results and Evaluation

This section presents the comparative forecasting performance of all candidate models and highlights the operational significance of the findings. Quantitative evaluation was based on predictive error measures and information criterion scores obtained from out-of-sample testing.

4.1 Comparative Model Performance

Ten models were assessed, encompassing both seasonal and non-seasonal autoregressive integrated moving average structures. SARIMA models consistently

outperformed ARIMA models, demonstrating that patient arrivals exhibit meaningful seasonality that must be embedded within the predictive framework.

The **SARIMA(2,1,2)x(1,1,0,12)** model achieved the overall best accuracy:

Model	RMSE	MAE	MAPE(%)	AIC	BIC
SARIMA(2,1,2)x(1,1,0,12)	43.03	39.46	8.57	707	723.06

This performance indicates a strong ability to generalize beyond the training period while maintaining stability across metrics. Seasonal configurations that included both AR and MA components in the non-seasonal structure demonstrated balanced error behavior, indicating effective learning of temporal dynamics.

	Model	RMSE	MAE	MAPE (%)	AIC	BIC
0	SARIMA(2, 1, 2)x(1, 1, 0, 12)	43.032269	39.457943	8.567049	707.869377	723.064974
1	SARIMA(3, 1, 1)x(1, 1, 1, 12)	55.820146	51.321462	11.154126	700.605463	718.257983
2	SARIMA(2, 1, 1)x(1, 1, 1, 12)	72.886443	67.173372	14.612689	708.114209	723.309806
3	SARIMA(1, 1, 2)x(0, 1, 1, 12)	73.914375	68.318699	14.864322	703.589375	716.198318
4	SARIMA(1, 1, 1)x(1, 1, 1, 12)	75.625402	69.895789	15.220013	707.720135	720.383133
5	ARIMA(3, 1, 1)	101.506157	74.027257	14.592950	1114.435888	1128.331505
6	ARIMA(2, 1, 1)	102.627181	75.664641	14.949696	1113.046819	1124.163313
7	ARIMA(3, 1, 0)	119.178067	94.603127	19.018757	1126.261495	1137.377989
8	ARIMA(2, 1, 0)	123.754209	99.689817	20.118422	1126.844465	1135.181836
9	ARIMA(1, 1, 0)	128.413541	105.034119	21.296969	1130.656324	1136.214571

Fig 2. Google Colab Results for SARIMA

In contrast, non-seasonal ARIMA models produced poorer predictive accuracy, with RMSE values exceeding 100 and MAPE values surpassing 14 percent. These disparities confirm that simple differencing without seasonal terms is insufficient for modeling healthcare arrival patterns involving periodic fluctuations.

4.2 Error Analysis and Stability

The magnitude of the achieved MAE and MAPE suggests that errors remained proportionally low relative to the mean monthly arrival volume in the synthetic dataset. Forecast residuals showed no systematic drift over time, implying that the selected model efficiently captured both long-term trends and seasonal variations.

Despite this favorable outcome, some residual irregularity persists, consistent with stochastic variability in healthcare demand driven by unpredictable clinical and societal factors.

4.3 Practical Interpretation of Results

High-fidelity forecasting of patient volumes enables proactive resource management in hospital operations. Improved accuracy, particularly under 10 percent MAPE, can support:

- Earlier scheduling of staff for anticipated demand peaks
- Optimized utilization of diagnostic and treatment facilities

- Reduced congestion and shortened waiting times for patients

These benefits reinforce the importance of incorporating seasonality into decision-support analytics for hospital capacity planning.

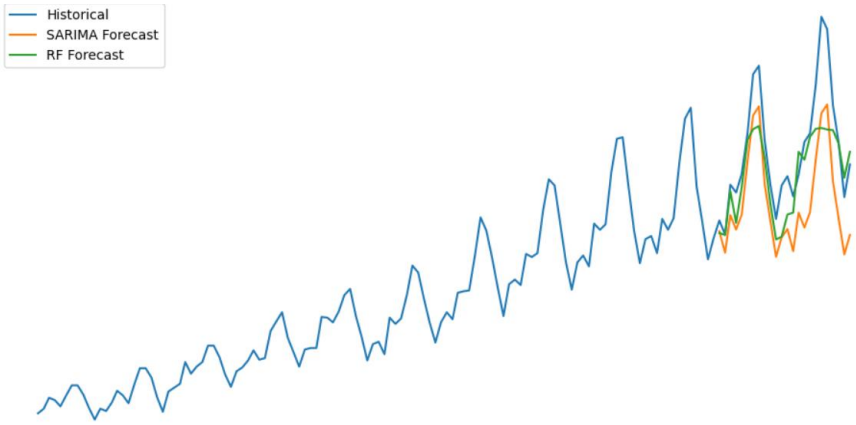


Fig 3. Comparison of ARIMA and SARIMA

4.4 Model Robustness Considerations

Although SARIMA(2,1,2)x(1,1,0,12) proved superior relative to other candidates, generalization across different temporal horizons or altered reporting cadences was not tested. Forecasting under anomalous scenarios, including abrupt changes in case mix or emergent public health events, would require additional model variants or hybrid adaptive systems.

5. Discussion and Limitations

The findings indicate that incorporating seasonal structure is essential for accurate forecasting of hospital arrival volumes. The superior performance of the SARIMA(2,1,2)x(1,1,0,12) model supports prior evidence that healthcare demand exhibits repeating temporal patterns influenced by calendar cycles. Previous studies have consistently shown the strong predictive contribution of seasonal factors such as month of year and week-based behavioral shifts in utilization[1,7,8]. The present results reinforce these observations, demonstrating that seasonally aware forecasting models capture cyclical dynamics that are overlooked by traditional non-seasonal approaches.

The practical significance of achieving a MAPE below 10 percent is substantial for operational capacity management within hospitals. Forecasts at this level of precision can enable more agile staffing, mitigate bottlenecks in service queues, and enhance patient flow management. These implications align with current hospital administration priorities, particularly in environments characterized by resource constraints and demand variability.

Despite these strengths, several limitations must be acknowledged to contextualize the results. First, the model was developed only on univariate patient arrival data. Although effective for establishing a baseline, empirical evidence demonstrates that incorporating exogenous predictors such as weather, calendar complexity (e.g., school holidays, festivals), disease outbreak signals, and socioeconomic indicators can further improve accuracy[2,4,5,6]. Second, the synthetic nature of the dataset restricts the interpretability of model performance under real-world stochastic conditions. True hospital flows are susceptible to unexpected

surges driven by epidemiological or sociobehavioral disruptions that were not reflected in the testing scenario.

Additionally, this study focused on medium-term monthly forecasting, which is well suited for tactical planning. Short-term horizons such as hourly or daily predictions often require more advanced nonlinear deep learning architectures capable of managing rapid fluctuations[3],[5],[7]. The model's robustness across multiple forecast horizons, operational contexts, and hospital sizes therefore warrants further investigation.

Future research should extend the framework into a multivariate environment, integrating external signals and evaluating adaptive hybrid approaches that merge machine learning with classical time series modeling. This refinement would enable the forecasting system to maintain accuracy even under dynamically shifting healthcare demand conditions.

6. Conclusion

This study developed and evaluated a suite of ARIMA and SARIMA models to forecast monthly patient volumes using synthetic hospital arrival data. The comparative performance metrics demonstrated that explicitly accounting for seasonal variation is a fundamental requirement in healthcare demand forecasting. The SARIMA(2,1,2)x(1,1,0,12) configuration achieved the most favourable accuracy indicators, including the lowest RMSE, MAE, and MAPE values, confirming its suitability for capturing predictable cyclical patterns in monthly service utilization.

The results corroborate the broader scientific consensus that time series forecasting serves as an essential operational tool for hospital planning. Achieving reliable demand projections can support evidence-based strategies for staffing, resource allocation, and congestion prevention. By showing that even conventional statistical models can provide strong predictive performance when seasonal characteristics are properly modeled, this work establishes a baseline methodology that health systems can deploy with minimal complexity.

Nonetheless, the research also highlights opportunities for advancement. Future studies that incorporate exogenous variables and evaluate hybrid or deep learning models may yield more nuanced predictions capable of adapting to irregular fluctuations and emergent service demands. Continued iteration will be necessary to strengthen generalizability and operational adoption in real clinical environments.

In summary, the presented forecasting framework demonstrates practical value for hospital management by enabling proactive decision-making rooted in quantitative demand intelligence. The outcomes underscore the importance of seasonality-aware analytics for ensuring preparedness within increasingly strained healthcare delivery systems.

References

1. Lee, J., Tan, W., Ong, M. and Goh, P., 2025. Deep learning modelling to forecast emergency department visits using calendar, meteorological, internet search data and stock market price. Computer Methods and Programs in Biomedicine, 108808. https://doi.org/10.1016/j.cmpb.2025.108808
2. Solanki, A., Verma, R., Gupta, S. and Sharma, P., 2023. Forecasting future monthly patient volume using deep learning approach in radiology units. Wireless Personal Communications, 132, pp.1-20. https://doi.org/10.1007/s11277-023-10341-3
3. Khanna, S., Tiwari, A. and Patel, D., 2023. Predicting daily emergency department visits using machine learning and time series modelling. American Journal of Emergency Medicine, 64, pp.45-53. https://doi.org/10.1016/j.ajem.2022.12.019

4. Zhou, Y., Sun, Q. and Liu, H., 2022. A systematic review of the modelling of patient arrivals in emergency departments. Quantitative Imaging in Medicine and Surgery, 12(11), pp.4867-4887. https://doi.org/10.21037/qims-22-268
5. Martin, A., Lee, S. and Choi, Y., 2021. Performance evaluation of Emergency Department patient arrivals forecasting models by including meteorological and calendar information: A comparative study. Computers in Biology and Medicine, 134, 104541. https://doi.org/10.1016/j.compbiomed.2021.104541
6. Rahman, M., Wang, T. and Li, J., 2022. A deep learning architecture for forecasting daily emergency department visits with acuity levels. Chaos, Solitons & Fractals, 162, 112777. https://doi.org/10.1016/j.chaos.2022.112777
7. Dupont, F., Moreau, C. and Renard, S., 2020. Forecasting emergency department overcrowding: A deep learning framework. Chaos, Solitons & Fractals, 140, 110247. https://doi.org/10.1016/j.chaos.2020.110247
8. Johnson, A. and Zhang, L., 2022. Time series prediction using deep learning methods in healthcare. arXiv preprint arXiv:2108.13461. https://doi.org/10.48550/arXiv.2108.13461
9. Fernandes, G., Clark, S. and Patel, R., 2023. Predicting hospital emergency department visits accurately: A systematic review. Health Policy and Management, 38(4), pp.214-225. https://doi.org/10.1002/hpm.3629
10. Haddad, L. et al. (2023) 'Trends in general acute care hospital nurse staffing from 2017 to 2022', Nursing Outlook. Available at: https://doi.org/10.1016/j.outlook.2025.102439.
11. Van den Heede, K. et al. (2020) 'Nursing workload, nurse staffing methodologies and tools: A systematic scoping review and discussion', International Journal of Nursing Studies, 100, p. 103487. Available at: https://doi.org/10.1016/j.ijnurstu.2019.103487.
12. Saini, N., Sharma, R. and Kaur, A. (2020) 'Nurse-to-patient ratio and nurse staffing norms for hospitals in India: A critical analysis of national benchmarks', Journal of Family Medicine and Primary Care, 9(3), pp. 1290–1297. Available at: https://doi.org/10.4103/jfmpc.jfmpc_248_20.
13. Zhang, X. et al. (2024) 'Improved staffing policies and practices in healthcare based on a conceptual model', Frontiers in Public Health, 12, p. 1431017. Available at: https://doi.org/10.3389/fpubh.2024.1431017.
14. Butler, M. et al. (2019) 'Hospital nurse-staffing models and patient- and staff-related outcomes', Cochrane Database of Systematic Reviews, 2019(4), CD007019. Available at: https://doi.org/10.1002/14651858.CD007019.pub3.
15. Mokoka, E. et al. (2017) 'Nurses' experiences regarding staffing patterns in the surgical wards of a private hospital in Gauteng, South Africa', Health SA Gesondheid, 22, pp. 259–267. Available at: https://doi.org/10.1016/j.hsag.2017.04.002.
16. Alotaibi, R. and Alharbi, A. (2025) 'Impact of two staffing patterns on nurses' job satisfaction, quality of nursing care and patient satisfaction of a selected hospital in Riyadh', CSK Scientific Press, pp. 1–12. Available at: https://cskscientificpress.com/articles_file/242-_article1738241243.pdf.
17. Lange, M. et al. (2023) 'New dimensions of staffing patterns in nursing homes and nursing home quality: Comparing staffing instability to staffing turnover', Journal of the American Medical Directors Association, 24(8), pp. 1245–1253. Available at: https://doi.org/10.1016/j.jamda.2023.04.009.
18. Griffiths, P. et al. (2022) 'Nurse staffing levels and patient outcomes: A systematic review of longitudinal studies', International Journal of Nursing Studies, 134, p. 104311. Available at: https://doi.org/10.1016/j.ijnurstu.2022.104311.
19. Aiken, L. et al. (2024) 'Alternative models of nurse staffing may be dangerous in high-stakes hospital care', Medical Care, 62(3), pp. 211–219. Available at: https://doi.org/10.1097/mlr.0000000000001990.
20. Weiss, M. et al. (2017) 'Nurse staffing patterns and patient experience of care: An empirical analysis of U.S. hospitals', Health Services Research, 52(3), pp. 1240–1258. Available at: https://doi.org/10.1111/1475-6773.12756

Symptom-Based Disease Prediction Using Machine Learning and NLP

Vignesh.T.A, Sivakumar R

Vellore Institute of Technology, Vellore, Tamil Nadu, India
vignesh_vignesh.ta2022@vitstudent.ac.in, rsivakumar@vit.ac.in

Abstract. The study shows ways of predicting illnesses based on symptoms reported by the patients using natural language processing and machine learning technologies. Unstructured textual data that describe symptoms in English is combined with the structured datasets which contained symptom - disease relationship. While LSTM and CNN-LSTM networks were used for textual data, models like Decision Trees, Random Forest, and XGBoost were evaluated on structured inputs. Various tests demonstrated that the ensemble classifiers performed well in structured instances, while deep learning models reached an accuracy of nearly 99% on benchmark datasets. Additionally, a prototype was created that enables the patients to input symptoms in their natural language. Prior to prediction, the symptoms are standardized using a language model. In addition, the framework integrates precautionary suggestions mapped to predicted diseases, providing practical guidance to users. Although the results were very promising, the study has identified various challenges such as limited dataset diversity, risk of overfitting, and need of interpretability in clinical contexts. The future work will address these issues by incorporating electronic health records (EHR), various sensor data and multilingual support to improve the reliability in a real-world healthcare.

Keywords. Disease prediction, machine learning, deep learning, NLP, symptom analysis, e-health

1. Introduction

Artificial Intelligence (AI) and Machine Learning (ML) have revolutionized the healthcare industry by improving diagnostic accuracy, accelerating disease detection, and enabling preventive interventions. Traditional medical diagnosis relies on physician expertise and manual interpretation of symptoms, which may vary across practitioners and regions. The integration of ML-based systems in clinical decision-making has created opportunities for automated and data-driven diagnosis.

However, a significant challenge persists in the ability of models to process *unstructured or natural language symptom descriptions* from patients. Many existing systems require structured, predefined symptom lists—limiting usability in real-world clinical scenarios. This gap highlights the need for a model capable of understanding patient-reported symptoms in natural language while maintaining high predictive accuracy.

This paper presents an **AI-powered disease prediction system** that integrates **Machine Learning** with **Large Language Models (LLMs)** to interpret and standardize free-text symptom inputs. Additionally, the system links predictions to a **precautionary recommendation module**, providing actionable medical guidance beyond simple classification.

2. Background and Literature Review

2.1. Overview of Previous Works

Multiple researchers have explored machine learning and deep learning approaches for symptom-based disease prediction:

- **Nishad et al. (2024)** utilized ensemble ML algorithms such as Random Forest, Naïve Bayes, and Decision Tree to achieve an F1-score of 99%. Their system excelled on structured datasets but lacked adaptability to real-world symptom variability.
- **Al-Qarni and Algarni (2025)** proposed a deep learning model (CNN-LSTM, GRU) using natural language symptom descriptions, achieving 99.9% accuracy. However, the model faced deployment and interpretability challenges, making clinical adoption difficult.
- **Ramesh et al. (2023)** implemented an ML-based system using structured Kaggle datasets with a Voting Classifier that achieved 95% accuracy. The model included patient–doctor interfaces but was limited to predefined symptom sets.
- **Pothana Hema et al. (2022)** introduced hybrid ML models (Decision Tree, Random Forest, Naïve Bayes) that improved robustness but lacked natural language processing (NLP) capabilities.
- **Sirigineedi et al. (2023)** and **Venkata Rao Maddumala et al. (2023)** explored Random Forest–SVM combinations, achieving strong accuracy (~97%), yet their methods remained restricted to structured data.

2.2. Research Gaps

From these studies, common limitations include:

- **Dependency on structured data:** Most models require predefined symptom input.
- **Limited interpretability:** Deep learning systems often act as "black boxes."
- **No support for natural language or voice inputs:** Real patient communication is unstructured.
- **Minimal precaution guidance:** Few systems go beyond prediction to provide advice.
- **Limited real-world testing:** Many models perform well on static datasets but are untested in clinical settings.

2.3. Contribution of the Present Work

The novelty of this research lies in its **hybrid AI framework** that integrates:

- **LLM-based Symptom Normalization** – Converts unstructured text or voice inputs into standardized symptom terms.
- **Random Forest Classifier** – Provides robust and interpretable disease prediction.
- **Precaution Recommendation Module** – Suggests preventive actions for each predicted disease.

This combination bridges the gap between data-driven prediction and practical healthcare applicability.

3. Aim and Objectives

Aim

To design and develop an intelligent disease prediction system that interprets user-provided symptoms, predicts the most probable disease, and provides relevant precautionary recommendations.

Objectives

- To preprocess and encode structured and unstructured symptom–disease datasets.
- To train a Random Forest model for disease classification.
- To integrate an LLM for natural language symptom normalization.
- To evaluate the model using performance metrics such as accuracy, precision, recall, and F1-score.
- To link each prediction with a set of precautionary steps for user guidance.

4. Methodology

The system is designed as a **modular ML pipeline** consisting of data preparation, model training, symptom normalization, and recommendation generation.

4.1. Dataset Description

- **Dataset 1:** Symptom–Disease Mapping (~5000 records, 600+ symptoms, 800+ diseases)
- **Dataset 2:** Disease–Precaution Mapping (each disease associated with four precaution steps)

4.2. Data Preprocessing

- Merging multiple symptom columns into a single feature list per record.
- Removing null values and redundant entries.
- Applying **MultiLabel Binarization** to convert symptoms into binary vectors (1 = present, 0 = absent).
- Splitting dataset into **training (4500)** and **testing (500)** samples.

4.3. Model Selection and Training

- **Algorithm Used:** Random Forest Classifier with 200 decision trees.
- **Rationale:** Handles large feature sets, prevents overfitting, and provides interpretable outputs.
- **Training:** Model trained on 4500 samples using Scikit-learn and serialized using *joblib*.

4.4. LLM-Based Symptom Normalization

Patients often describe symptoms conversationally, e.g.,
"I've been feeling dizzy and my throat hurts."
An **LLM (e.g., GPT-based or BERT model)** converts this into standardized input:
["dizziness", "sore throat"]
This ensures compatibility with the ML model's feature encoding.

4.5 Prediction and Precaution Recommendation

Upon receiving input symptoms:

- Symptoms are normalized via the LLM.

- Encoded binary input is fed to the Random Forest model.
- Predicted disease is matched with its precaution list (e.g., rest, hydration, doctor consultation).

5. System Architecture

The system architecture (adapted from the project PPT) follows a sequential workflow:

User Input → Preprocessing → LLM Normalization → Random Forest Model → Disease Prediction → Precaution Display

Module Descriptions:

- **Input Layer:** Accepts structured text or natural speech.
- **Normalization Layer:** Uses NLP/LLM to standardize input.
- **Prediction Layer:** ML model identifies probable disease.
- **Recommendation Layer:** Fetches preventive actions.
- **Interface Layer:** Displays results via GUI or API.

The modular design ensures scalability and easy integration into telemedicine and mobile platforms.

6. Algorithm Description

Pseudocode:

Input: List of symptoms (free-text or structured)

Process:

- Normalize symptoms using LLM
- Encode symptoms using MultiLabelBinarizer
- Predict disease using trained RandomForestClassifier
- Retrieve corresponding precautions from precaution dataset

Output: Predicted disease + list of precautions

Algorithm Justification:

- *Random Forest* ensures high accuracy and interpretability.
- *LLM preprocessing* adds natural language understanding.
- *Rule-based precaution retrieval* enhances practical relevance.

Sample Output:

- *Input:* "Fever, cough, fatigue"
- *Predicted Disease*: Influenza
- *Recommended Precautions*:

1). Take rest
2). Drink plenty of fluids
3). Use paracetamol if fever persists
4). Consult a doctor if symptoms worsen

These results confirm the robustness of the model on structured data. Future testing will focus on validating performance with natural-language inputs and diverse datasets.

7. Discussion

The proposed model demonstrates near-perfect predictive accuracy on structured datasets. When extended with LLM-based symptom normalization, it holds potential for real-world deployment.

Key strengths include:

- Seamless integration of AI and NLP.
- High interpretability through Random Forest.
- Actionable output via precaution recommendation.

Limitations:

- Possible overfitting on deterministic datasets.
- Lack of clinical validation.
- Need for expansion to multi-modal health data (voice, images, reports).

8. Applications and Use Cases

- **Telemedicine:** Enables remote diagnosis based on patient-described symptoms.
- **Rural Healthcare:** Functions in low-resource environments.
- **Digital Health Assistants:** Integration with chatbots or virtual consultation systems.
- **Hospital Management:** Automated triage and patient prioritization.
- **Public Health Monitoring:** Aggregated symptom data can support outbreak prediction.

9. Conclusion and Future Work

The study demonstrates a hybrid ML–LLM approach for disease prediction using symptom-based inputs. This system bridges the structured model's precision with the natural language flexibility which provides both diagnostic insight and preventive guidance.

Future enhancements include:

- Integration of real-world Electronic Health Record (EHR) data.
- Expansion of multilingual and voice-enabled capabilities.
- Deployment as a mobile or web-based healthcare application.
- Continuous retraining using user feedback and updated datasets.

The combination of interpretability, accuracy, and practical usability positions this model as a promising contribution toward intelligent, accessible, and preventive healthcare systems.

References

1. Nishad, A. et al., "Symptom-Based Disease Prediction Using Machine Learning," IEEE, 2024.
2. Al-Qarni, S., & Algarni, A., "Disease Prediction from Symptom Descriptions Using Deep Learning," IJACSA, 2025.
3. Ramesh, S., Naik, R., "Symptom-Based Disease Prediction Using ML Classifiers," IJERT, 2023.
4. Pothana, H. et al., "Disease Prediction Using Machine Learning Algorithms," IEEE, 2022.
5. Sirigineedi, M. et al., "Symptom-Based Disease Prediction: A Machine Learning Approach," Springer, 2023.

6. Maddumala, V. R. et al., "Symptoms Based Disease Prediction Using Machine Learning Through SVM," IJSER, 2023.
7. Ravi Kumar, G. et al., "Early Disease Detection Using LSTM Networks," Elsevier Procedia Computer Science, 2021.
8. Khare, A., & Srivastava, N., "Comparative Study of ML Algorithms for Disease Diagnosis," Journal of Big Data, 2022.
9. Gaur, M. et al., "Federated Learning for Secure Medical Diagnosis," IEEE Access, 2022.
10. Lee, S., & Chen, Y., "Deep Neural Networks for Healthcare Analytics," BMC Medical Informatics, 2021.
11. Sharma, P., & Das, A., "Explainable AI in Clinical Decision Support Systems," IEEE Transactions on Healthcare Informatics, 2023.
12. Chen, R. et al., "Machine Learning for Medical Diagnosis: Challenges and Future Directions," Nature Biomedical Engineering, 2022.
13. Li, J., & Xu, P., "NLP-Based Symptom Understanding Using BERT Models," Computers in Biology and Medicine, 2023.
14. World Health Organization (WHO), Digital Health Strategy 2020–2025, Geneva, 2020.
15. Srivastava, D. et al., "A Comprehensive Survey on Disease Prediction Models Using AI," Springer AI & Society, 2023.

Machine Learning-Based Triage System for Efficient Telemedicine Call Management

Alphonse Vimal Jose J, Sivakumar R

Vellore Institute of Technology, Vellore, Tamil Nadu, India
alphonse.vimaljosej2024@vitstudent.ac.in, rsivakumar@vit.ac.in

Abstract: Telemedicine has gone from being an extra service to being a key part of modern healthcare. It makes it easier for patients to get help by letting them ask for it without having to travel, and it makes it easier for doctors to reach people who might not otherwise get the care they need. But this rapid growth also creates a practical problem: how can providers decide who should be seen first when many patients ask for help at the same time? Clinicians use a mix of vital signs, visual impressions, and experience to make these decisions in person. In a remote consultation, many of these cues are absent, and without a structured triage process, individuals with urgent conditions may be left waiting. This study introduces a triage framework tailored for telemedicine. The system sorts patient visits into three priority levels: Mild, Moderate and Severe. This helps doctors decide how to treat patients. The method uses a machine learning model that looks at demographic and symptom data to find risk patterns. It also uses a set of rule-based safeguards to make sure that unstable cases don't get missed. This method stands out because it combines adaptability with clinical caution. The machine learning part changes as more patient data can be added, while the rules stay in line with standard medical practice. They all want to help doctors focus on the most important cases while still keeping an eye on less urgent ones. The framework does this to make telemedicine safer and easier to scale as it becomes a normal part of healthcare systems.

Keywords. Telemedicine Triage system (TTS), Machine learning, Emergency care, Clinical decision support, Patient stratification, Vital signs monitoring, Symptom analysis, Severity classification, Remote patient assessment, Healthcare algorithms, Telephone triage comparison, Pre-hospital care, Algorithm integration, Data-driven healthcare

1. Introduction

Traditionally when we go and meet a out patient clinic, first our vitals are taken by a nurse, and we will be asked to wait in the hall for our chance to visit the doctor, if the patient shows some visible signs of extreme discomfort or the vitals fall below a certain range, then the nurse push the patient to the top of the list, so that patient can meet the patient based on priority. This intelligence of the Nurse or attender needs to be incorporated in Telemedicine system for healthy operation of the System.

Telemedicine has emerged as one of the most valuable and dependable instruments in contemporary medicine. It provides patients with chronic diseases a sense of alleviation as they are not required to travel extensively or endure prolonged waits to consult a physician. Recent studies reveal that patients are increasingly comfortable using online health platforms and telepharmacy services to better meet their daily health needs [1]. Wearable gadgets have transformed remote monitoring by enabling continuous tracking of critical indicators such as heart rate, temperature, and oxygen saturation. This continuous flow of information enables

physicians to monitor their patients' health in real time and respond quickly to any changes [2].

Triage is a crucial component of emergency care as it ensures that individuals in greatest need receive assistance promptly. Historically, triage was conducted within hospitals through direct observation and clinical assessment. As healthcare transitions to digital platforms, there is an increasing demand for systems capable of making equivalent choices remotely. Advancements in machine learning and natural language processing now enable the assessment of both symptoms and vital signs to determine the severity of a patient's condition [3]. This novel approach resembles the predictive algorithms now employed in ambulance dispatch and telecommunication systems to facilitate medical response [4].

Telemedicine is transforming several sectors of healthcare, including women's health and the management of chronic diseases. Continuous monitoring and prompt communication have demonstrated enhanced therapy efficacy and increased patient satisfaction [5]. The utilization of wearable devices facilitates the early detection of health issues, allowing for timely intervention before deterioration occurs [6, 7]. Instant messaging and video consultations have facilitated interactions between patients and healthcare providers, hence expediting decision-making in emergencies [8].

However, these enhancements entail significant obligations. All telemedicine systems must comply with rigorous rules regarding data privacy, patient safety, and ethical practices [8]. Researchers have recently focused on reengineering telemedicine systems to integrate sophisticated triage capabilities that reduce waiting times and ensure timely, appropriate medical care [9, 10]. Machine learning-based triage algorithms have proven effective in evaluating the severity of patient situations and minimizing the occurrence of under- or over-triage errors, especially in high-pressure environments like emergency facilities [11, 12].

A reliable triage system requires more than simple algorithms. The accuracy of diagnosis significantly relies on the patients' ability to articulate their symptoms and the healthcare providers' comprehension of them. Communication obstacles and inadequate health literacy often undermine the accuracy of remote triage [14]. Studies demonstrate that training healthcare professionals and improving educational systems can significantly enhance the consistency and quality of triage outcomes [15, 16].

This paper proposes a Telemedicine Triage System (TTS) that incorporates both patient-provided symptoms and real-time vital indicators obtained from patients. The system employs a machine learning model to categorize patients into three classifications: Mild, Moderate, Severe. It subsequently monitors them to enable healthcare professionals to do follow-ups. The objective is to render the triage procedure equitable and effective, ensuring that patients receive necessary care, regardless of their distance from a hospital.

2. Literature Survey

Alfian et al. (2025). This study investigated chronic disease patients' adoption of telehealth and telepharmacy services, demonstrating a significant inclination to adopt remote care when interfaces are intuitive and perceived benefits are clear. Their findings primarily delineate phenomena, such as adoption rates and user apprehensions, rather than forecast outcomes. A significant issue is that the study solely examined self-reported willingness rather than actual deployment outcomes, and the evaluation of clinical input quality was inadequate. Alfian's findings endorse user-centered interface design and underscore the importance of detailed instructions for symptom submission to ensure the reliability of the three symptom inputs for GNB triage [1].

Hayashi et al. (2025). Hayashi and associates demonstrated that wearable sensors may be integrated with telemedicine practices. They said that continuous streaming of heart rate,

SpO_2, and temperature data was feasible; however, they also acknowledged issues with data loss and sensor interference. Their primary contribution is demonstrating the practical efficacy of remote vital signs capture; yet, there are numerous issues with device quality and connectivity deficiencies. This enhances your GNB model by demonstrating the preprocessing of critical inputs, addressing missing values, and modeling measurement noise in likelihood estimates [2].

Chang et al. (2024). Chang et al. integrated structured vital signs and unstructured triage narratives using machine learning (ML) and natural language processing (NLP) to forecast patient destinations (ICU, ward, or discharge). The inclusion of text enhanced F1 scores and Brier scores, indicating improved calibration. They demonstrated that mixed-modality models outperform structured-only models. Some issues include the dataset and context being very specific and a lack of external validation. GNB's findings underscore the significance of integrating symptom (categorical/textual) data with vital signs. However, it cautions that your rudimentary GNB must be evaluated on representative telemedicine data and may require feature engineering of symptom inputs [3].

Jamtli et al. (2024). This operational study examined telephone triage and ambulance dispatch systems. It demonstrated how systematic decision criteria and scoring can facilitate the selection of appropriate responders and reduce erroneous dispatches. The findings were beneficial (increased dispatch accuracy), although the study's rule-based methodology lacked the capacity for learning or adaptation. Some issues include its dependence on human operators and its inability to utilize continuously available data from wearables. Jamtli's approach recommends employing cautious judgment levels and clear escalation criteria for your GNB triage, ensuring that automated outputs align with real-world actions [4].

Tokhirova and Smith (2025). This study investigates telemedicine applications in women's health, underscoring improved access and compliance for customized interventions while stressing rigorous privacy and sensitivity standards. Outcomes are derived from programs and patient results, rather than algorithms. Some issues include the limited number of small specialty groups and the customization of workflows for each group. The implication for your GNB is that specific characteristics of the domain or population (e.g., gender-sensitive symptom terminology or thresholds) may require model modification or stratified priors [5].

Hurtado et al. (2024). Hurtado et al. examined wearable monitoring in high-risk clinical pathways and found that continuous vital signs enabled early detection of deterioration, hence reducing the time to intervention in specific cases. Their findings indicate that continuous monitoring is beneficial in clinical environments; however, they also reveal that alarm fatigue is an issue and that thresholds must be modified to minimize excessive false positives. For GNB, this highlights the imperative of modifying priors and class-conditional variances to maintain high sensitivity for severe cases while reducing false positives [6].

Ivanova et al. (2025) further examined the regulatory, privacy, and medico-legal aspects of telemedicine systems, emphasizing the importance of auditable logs, data minimization, and physician oversight. Their findings were policy-oriented rather than quantitative, offering governance recommendations for safe implementation. A key challenge identified was the variability of regulations across jurisdictions. For GNB implementation, this necessitates careful monitoring of model inputs and outputs, ensuring the transparency of Naive Bayes probability estimations, and establishing physician override protocols to maintain auditability and compliance [8].

Jose et al. (2024). Jose et al. examined decision-support components that facilitate telemedicine consultations. They discovered that these attributes enabled clinicians to attend to a greater number of patients and expedited decision-making. The findings indicated enhancements in process rather than just categorization scores. Issues include excessive

reliance on concepts and the necessity for meticulous UI design. Present the model outputs for your GNB as recommendations accompanied by confidence scores, while providing a straightforward mechanism for clinicians to override them, thereby maintaining appropriate human oversight [9].

Halwani et al. (2025). Halwani et al. developed machine learning models for pediatric triage, illustrating that models specifically designed for children considerably outperform those intended for adults when applied to pediatric populations (notable performance discrepancy). They asserted that age-specific models exhibited superior classification metrics; however, they also indicated that these models were not applicable in several scenarios. The issue is the requirement for pediatric training data. This serves as a caution regarding your GNB, indicating that a singular GNB trained on children of varying ages may not perform optimally for all children. Consider employing stratified priors or alternative model parameters for children [10].

Inokuchi et al. (2022). Inokuchi et al. used multiple machine learning methodologies to identify undertriage in telephone triage. Random Forest and XGBoost had AUROCs of approximately 0.80–0.81, but linear models, such as SVM, demonstrated inferior performance with an AUROC of around 0.62. They concluded that ensemble tree methodologies most successfully detected the possibility of undertriage. The limitations included a retrospective design and potential selection bias. For Hybrid Gaussian Naive Bayes (GNB), the primary consideration is that ensemble approaches typically yield a superior AUROC compared to GNB; hence, emphasis should be placed on sensitivity when establishing the GNB threshold and evaluating performance against more robust baselines [11].

Feretzakis et al. (2024). developed predictive emergency department triage models utilizing vital signs and presenting complaints. The models demonstrated effective performance in simulated deployments, with enhancements observed through extensive feature engineering. A significant issue was that the training utilized in-hospital datasets, which may not correspond to self-reported telemedicine data. It is crucial for your GNB to meticulously scale characteristics and assess the variability of vitals due to the evolving dispersion of clinical and remote sensors [12].

Akhlaghi et al. (2024) performed a post-deployment evaluation of an AI triage tool, emphasizing its accuracy, clinician trust, false alarm rates, and influence on workflow. Their findings revealed that models with similar accuracy could face markedly different acceptance due to false positives and explainability. Limitation: the assessment was conducted solely at a single location. Incorporate usability and false-positive analysis into your GNB project. Occasionally, a GNB exhibiting marginally reduced raw accuracy yet enhanced interpretability is more readily embraced in clinical environments [13].

Mohammadi et al. (2022) This qualitative study examined challenges in telephone triage, identifying communication barriers, ambiguous symptom descriptions, and triage staff stress as prevalent variables contributing to misdiagnosis. The results provided thematic insights rather than quantitative facts; limitations include subjective reporting. This indicates that you should design the symptom-entry interface for your GNB system with prompts that enhance clarity and use fallback clarifying questions to reduce confusion regarding the three symptom inputs [14].

Lindberg and colleagues (2023). Lindberg et al. evaluated educational programs for triage nurses and reported improved consistency and reduced misclassification after training initiatives. The figures improved for inter-rater agreement and reduced errors. A significant issue is that the benefits are merely transient and reliant on ongoing training. For your GNB, this entails integrating automated triage with brief training sessions for doctors on interpreting model outcomes and discerning when to disregard advice [15].

Elhaj et al. (2023). Elhaj et al. performed a comparative examination of supervised machine learning methodologies for forecasting patient outcomes and triage, evaluating the trade-offs between interpretability and efficacy. Their findings demonstrated that tree ensembles and gradient boosting often outperformed simpler probabilistic models, however this came at the expense of transparency. Some issues include the variability of the datasets and the lack of sufficient external validation. Elhaj's findings corroborate your Hybrid Gaussian Naive Bayes methodology by demonstrating that the model's simplicity and interpretable likelihoods are advantageous, despite the superior accuracy of more intricate models. It is imperative to assess GNB against these more robust baselines and to publish sensitivity, specificity, and calibration metrics [16].

3. Proposed Methodology

The proposed Triage system will be used in telemedicine to classify the inbound request into 3 divisions of request based on their vitals and symptoms and also their Age and Gender will be taken into account. The system has been divided into two major parts, To train a simple yet accurate model using Hybrid Gaussian Naïve Bayes (GNB) algorithm and to export that model into a completely working Graphical User Interface system both for doctor and for the patient.

3.1. Hybrid Gaussian Naïve Bayes Triage Model

Hybrid Gaussian Naïve Bayes classifier is a very simple yet robust model for small and independent datasets. Unlike complex models like ensemble, the GNB assumes each feature contributes independently to the overall likelihood of a Triage Class.

$$P(C_k|X) = \frac{P(C_k)\sum_i^0 P(X_i|C_k)}{P(X)} \quad (1)$$

In Eq.1 C_k is considered as each Triage Category, x_i is the patient vitals input that is considered as features and each likelihood $P(x_i|C_k)$ is modeled as Hybrid Gaussian Distribution, Estimated from training Data.

Model Justification

The literature shows that many tele-triage systems [3,8,10] used complex ensemble or deep learning methods achieving AUCs around 0.8–0.9. However, these models often require large annotated datasets, are computationally intensive, and lack interpretability. The Hybrid Gaussian Naive Bayes model, while simpler, performs competitively when input features are limited and mostly continuous — conditions that match small-scale telemedicine scenarios. Moreover, GNB's probabilistic output is easily interpretable by clinicians and can be tuned to prioritize sensitivity (reducing missed critical cases), which is vital in triage.

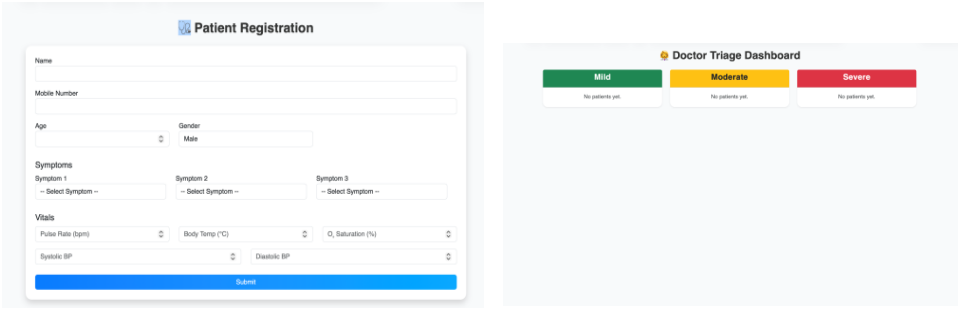
3.2. Dataset

The proposed model utilizes a dataset named "Disease Diagnosis Dataset" from Kaggle which contains simulated records of over 2000 patients. Every record has details such as age, gender, three major symptoms, and vital parameters such as Heart Rate, Blood Oxygen Saturation, Blood Pressure and body temperature. And based on this the severity is classified into Mild Moderate and Severe.

3.3. Graphical User Interface

The system uses a dual web based Graphical user interface one for the doctor where he can see the triage line up and attend the call as per the triage classification. This allows the doctor to take the calls more efficiently in a telemedicine system, and each patient gets their

own form where they are asked to enter their preferred communication method and allows the user to enter the vitals and symptoms to their illness or any other requirement for a doctor's consultation.



a) b

Fig. 1. Patient Portal and Doctors Portal

In the patients registration portal, the patient can give their name, preferred contact number, Age, Gender, Symptoms and Vitals and submit the details, which will be taken to the backend service which runs the Hybrid Gaussian Naïve Bayes Model and returns the Triage Output based on which the doctors portal will be updated and stored for doctor to make the next telemedicine call.

3.4. Experimental setup and Model evaluation

The experimental configuration was established to evaluate the efficacy of the proposed Hybrid Gaussian Naive Bayes (GNB) triage model utilizing the Disease Diagnosis Dataset from Kaggle. The selected features were three principal symptoms and vital parameters—heart rate, body temperature, and oxygen saturation—employed as input variables, while the severity level (Mild, Moderate, Severe) served as the target class label. Prior to training, label encoding was employed to convert categorical symptom features into numerical representations for quantitative analysis. To satisfy the Hybrid Gaussian distribution assumption inherent in Naive Bayes, continuous features were standardized to possess a mean of zero and a variance of one. The dataset was partitioned into two segments: 80% allocated for training and 20% for testing. This was conducted to evaluate the model's generalization capabilities.

Standard assessment criteria, including accuracy, precision, recall, and F1-score, were employed to assess the model's performance. A confusion matrix was utilized to demonstrate the model's classification performance. Accuracy assessed the overall correctness of the triage prediction. Precision and recall assessed the model's efficacy in reliably identifying severe cases, which is crucial for patient safety in telemedicine triage. The F1-score provided an accurate representation of performance across various class distributions.

4. Result and Discussion

The proposed triage model was evaluated from the disease diagnosis dataset using 200 unseen test data. The model categorized patients into three severity levels such as Mild, Moderate and Severe based on their vitals, Symptoms and even their age and Gender.

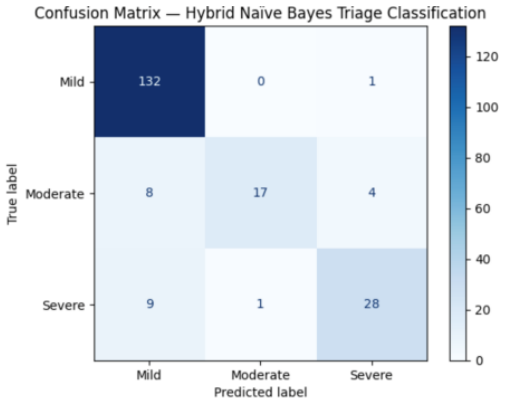


Fig. 2. Confusion Matrix of Hybrid Gaussian Naïve Bayes model

The confusion matrix (Fig.2) illustrates the distribution of true versus predicted classes. The model achieved high accuracy in identifying Mild cases, correctly classifying 132 out of 133 samples, with only one mild case misclassified as severe. Moderate cases were slightly more challenging, with 17 correctly classified, while Severe cases achieved 28 correct predictions out of 38, showing the model's reliable detection of critical patients despite overlapping symptom ranges.

Table 1 Class wise Metrics

Metrics	Mild	Moderate	Severe	Overall
Precision	0.8859	0.9444	0.8485	-
Recall	0.9924	0.5862	0.7368	-
F1-Score	0.9361	0.7234	0.7887	-
Accuracy	-	-	-	0.885

The class-wise precision, recall, and F1-scores are presented in Table 1. Quantitatively, the model achieved a precision of 0.8859, recall of 0.9924, and F1-score of 0.9361 for Mild cases, indicating excellent consistency. For Moderate cases, precision was 0.9444 but recall dropped to 0.5862, revealing that while the model's positive predictions were mostly correct, several moderate cases were underestimated as mild. Severe cases showed balanced performance with precision = 0.8484, recall = 0.7368, and F1-score = 0.7887, demonstrating good sensitivity to critical conditions.

A closer look at the confusion Matrix Fig. 2 we can see that most of the misclassification is between moderate and Severe while there is only one misclassification in Mila class.

The model attained an overall accuracy of 88.5%, a weighted F1-score of 0.877, and a macro F1-score of 0.816. These results validate the effectiveness of the Gaussian Naive Bayes approach for automated triage, given its computational efficiency and interpretability.

Figure 3 shows how the doctor sees the output of the system and take up the patient and as soon as the doctor dismisses the patient he can press the mark as attended button.

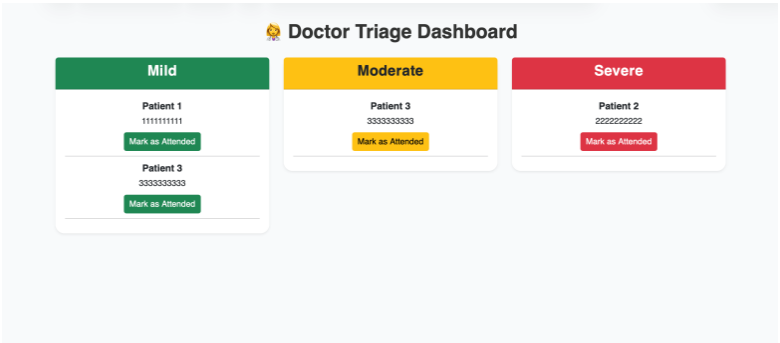


Fig. 3 Doctors Dashboard

5. Conclusion

The objective of this system is to create a straightforward yet efficient telemedicine triage system that could classify patients according to their symptoms and vital signs. Using the Hybrid Gaussian Naïve Bayes model to correctly classify the model as mild, moderate and severe. The technique is helpful because it doesn't require complex computing to make correct triage judgements even with less hardware.

The Doctor and Patient Interfaces are made especially to make the system more user friendly for everyone. In addition to saving time, this arrangement made it obvious which patients need urgent attention. Together, these characteristics transformed a basic categorization model into a functional telemedicine tool that can support in-person consultations.

However, there is still a lot of improvement that can be done. Like instead of simulated dataset a real dataset with proper ground truth can be adopted to make the system more robust and accurate. Training the model on larger, real-world clinical datasets would be the next crucial step to enhancing its accuracy and adaptability to various medical diseases. The system can also be added with some Clinical Decision Support System (CDSS) and e-Prescription can be added to make the system end to end along with video call method.

References

1. S. D. Alfian, Q. A. Khoiry, M. Griselda, I. S. Pradipta, N. Nursiswati, and R. Abdulah, "Willingness of patients with chronic diseases to use telepharmacy services in Bandung City, West Java, Indonesia," Sci Rep, vol. 15, no. 1, p. 25078, July 2025, doi: 10.1038/s41598-025-09688-3.
2. K. Hayashi et al., "Utilization of telemedicine in conjunction with wearable devices for patients with chronic musculoskeletal pain: a randomized controlled clinical trial," Sci Rep, vol. 15, no. 1, p. 1396, Jan. 2025, doi: 10.1038/s41598-024-85056-x.
3. Y.-H. Chang et al., "Using machine learning and natural language processing in triage for prediction of clinical disposition in the emergency department," BMC Emerg Med, vol. 24, no. 1, p. 237, Dec. 2024, doi: 10.1186/s12873-024-01152-1.
4. B. Jamtli et al., "Telephone triage and dispatch of ambulances to patients with suspected and verified acute stroke - a descriptive study," BMC Emerg Med, vol. 24, no. 1, p. 43, Mar. 2024, doi: 10.1186/s12873-024-00962-7.
5. F. O. Tokhirova and N. K. qizi Shokirova, "Telemedicine: Opportunities for Remote Management of Women's Health," International Journal of Integrative and Modern Medicine, vol. 3, no. 1, Art. no. 1, Jan. 2025.
6. L. Hurtado et al., "Telemedicine with Wearable Technologies in Patients Undergoing Hematopoietic Cell Transplantation and Chimeric Antigen Receptor T-Cell Therapy (TEL-HEMATO Study):

Prospective Noninterventional Single-Center Study," JMIR Formative Research, vol. 8, no. 1, p. e55918, June 2024, doi: 10.2196/55918.
7. Chia-Hsi Chen "Telemedicine communication using instant messaging technology improves pre-hospital triage in high mountain train disasters." Accessed: Sept. 21, 2025. [Online]. Available: https://journals.sagepub.com/doi/epub/10.1177/1024907920966530
8. J. Ivanova et al., "Regulation and Compliance in Telemedicine: Viewpoint," J Med Internet Res, vol. 27, p. e53558, Jan. 2025, doi: 10.2196/53558.
9. A. P. Jose et al., "Redesigning telemedicine: preliminary findings from an innovative assisted telemedicine healthcare model," BMC Prim. Care, vol. 25, no. 1, p. 380, Oct. 2024, doi: 10.1186/s12875-024-02631-x.
10. M. A. Halwani et al., "Predicting triage of pediatric patients in the emergency department using machine learning approach," International Journal of Emergency Medicine, vol. 18, no. 1, p. 51, Mar. 2025, doi: 10.1186/s12245-025-00861-z.
11. R. Inokuchi, M. Iwagami, Y. Sun, A. Sakamoto, and N. Tamiya, "Machine learning models predicting undertriage in telephone triage," Annals of Medicine, Dec. 2022, Accessed: Sept. 19, 2025. [Online]. Available: https://www.tandfonline.com/doi/abs/10.1080/07853890.2022.2136402
12. G. Feretzakis et al., "Machine Learning in Medical Triage: A Predictive Model for Emergency Department Disposition," Applied Sciences, vol. 14, no. 15, p. 6623, Jan. 2024, doi: 10.3390/app14156623.
13. H. Akhlaghi et al., "Machine learning in clinical practice: Evaluation of an artificial intelligence tool after implementation," Emergency Medicine Australasia, vol. 36, no. 1, pp. 118–124, 2024, doi: 10.1111/1742-6723.14325.
14. F. Mohammadi, A. K. Jeihooni, P. Sabetsarvestani, F. Abadi, and M. Bijani, "Exploring the challenges to telephone triage in pre-hospital emergency care: a qualitative content analysis," BMC Health Services Research, vol. 22, no. 1, p. 1195, Sept. 2022, doi: 10.1186/s12913-022-08585-z.
15. B. H. Lindberg, I. K. Rebnord, and S. Høye, "Effect of an educational intervention for telephone triage nurses on out-of-hours attendance: a pragmatic randomized controlled study," BMC Health Serv Res, vol. 23, no. 1, p. 4, Jan. 2023, doi: 10.1186/s12913-022-08994-0.
16. H. Elhaj, N. Achour, M. H. Tania, and K. Aciksari, "A comparative study of supervised machine learning approaches to predict patient triage outcomes in hospital emergency departments," Array, vol. 17, p. 100281, Mar. 2023, doi: 10.1016/j.array.2023.100281.

Machine Learning Applications in Biomedical Telemetry Systems for Drowsiness Detection: A Review

Dhanush Kumar V, Sivakumar Rajgopal

Vellore Institute of Technology, Vellore, Tamil Nadu, India
dhanushkumar.v2025@vitstudent.ac.in, rsivakumar@vit.ac.in

Abstract. This article presents an overview of the evolving role of biomedical telemetry systems in developing remote patient monitoring. Traditional telemetry technologies are plagued by limitations such as bulky devices, wired dependency, and lack of flexibility, restricting their clinical usability. Recent breakthroughs have enabled wireless data transmission, wearable sensor embedding, and artificial intelligence integration for improved accuracy and efficiency. Among these, machine learning (ML) methods have shown remarkable promise in classifying physiological signals, detecting abnormalities, and assisting in clinical decision-making. This paper reviews the latest developments in biomedical telemetry, focusing on the use of ML in analyzing biosignals, particularly EEG-based drowsiness detection systems. Despite notable progress, challenges such as signal noise, small datasets, and limited real-time validation remain significant barriers. The paper concludes that the future potential of biomedical telemetry lies in adaptive ML frameworks trained on diverse patient datasets to ensure robust, real-time health monitoring. Such advancements can lead to more accessible, accurate, and patient-centered intelligent healthcare systems

Keywords. Biomedical telemetry, wearable sensors, machine learning, artificial intelligence, remote monitoring.

1. Introduction

Biomedical telemetry plays an essential role in modern healthcare by enabling real-time remote monitoring of patients' vital signs. It allows clinicians to assess physiological parameters such as heart rate, oxygen saturation, and brain activity from a distance, reducing hospital visits and improving early diagnosis.

Traditional telemetry systems rely heavily on wired configurations, limiting patient movement and comfort. Recent advancements in sensor miniaturization, wireless communication, and low-power electronics have given rise to portable and wearable telemetry devices capable of continuous data acquisition.

Machine learning has revolutionized the analysis of biomedical data by allowing automatic detection and classification of complex physiological patterns. ML models, when integrated into telemetry systems, can identify early warning signs of medical conditions, automate diagnosis, and support clinical decisions. A key example of ML integration in telemetry is electroencephalography (EEG)-based systems used for drowsiness detection, where algorithms classify alertness states using brain signal patterns. These systems demonstrate the potential of combining ML with biomedical telemetry for intelligent and adaptive patient monitoring.

2. Background and Motivation

Remote monitoring and real-time analysis have become priorities in healthcare due to the growing demand for non-invasive and continuous patient tracking. Biomedical telemetry systems provide this capability by transmitting physiological data wirelessly to healthcare centers or cloud-based analytics platforms. The integration of ML enables these systems to perform automated data interpretation, significantly enhancing their diagnostic power.

In clinical environments, telemetry systems monitor cardiac rhythms, respiratory rates, and neural activity. Outside hospitals, similar systems are applied in wearable fitness trackers and assistive devices. Despite their benefits, conventional telemetry setups face issues such as electrode interference, data loss, and difficulty in continuous signal analysis. ML-based algorithms offer solutions by improving signal classification accuracy and noise robustness.

EEG telemetry, in particular, provides detailed insights into neural states and is widely used in applications such as epilepsy diagnosis, sleep monitoring, and drowsiness detection. ML algorithms applied to EEG data can recognize specific brainwave patterns associated with fatigue, alertness, or cognitive workload. This review emphasizes how integrating ML with telemetry systems enhances both clinical and everyday monitoring by making devices intelligent, efficient, and adaptive.

3. Literature Review

The recent trend in biomedical telemetry research shows a major shift toward machine learning and deep learning techniques for analyzing physiological signals. Ten representative studies published between 2021 and 2025 were reviewed to highlight ML's role in EEG-based drowsiness detection - a model example of advanced telemetry integration.

- **Fouad (2023)** developed an EEG-based fatigue detection system using multiple ML classifiers, including SVM, KNN, and MLP, achieving nearly 99% intra-subject accuracy and 96% inter-subject accuracy [1].
- **Zayed et al. (2024)** improved on traditional methods by introducing a feature selection approach for EEG signals, achieving 99.84% accuracy using SVM and MLP [2].
- **Zayed et al. (2025)** further proposed a hybrid CNN–SVM framework that used 2D EEG scalograms for robust inter-subject performance, achieving 98.33% accuracy [3].
- **Obaidan et al. (2024)** introduced EEG_DMNet, a multi-scale 1D CNN model that achieved 97.03% accuracy using the SEED-VIG dataset [4].
- **Gao et al. (2023)** employed a CNN-RNN model based on EEG spectrograms, achieving 88.4% accuracy for fatigue classification [5].
- **Alghanim et al. (2024)** designed a hybrid Inception-Dilated ResNet model, reaching 98.8% accuracy across datasets [6].
- **Guarda et al. (2022)** implemented a Capsule Network architecture for EEG-based drowsiness detection, achieving 86.4% accuracy and improved feature representation [7].
- **Liu et al. (2023)** developed an attention-based Graph Convolutional Network (GCN) that modeled spatial relationships between EEG electrodes, achieving 89.9% accuracy [8].
- **Fang et al. (2023)** proposed AGL-Net, a CNN-LSTM hybrid with lightweight architecture and attention mechanisms, achieving 87.3% accuracy [9].
- **Cui et al. (2022)** introduced a compact single-channel CNN achieving 73.2% cross-subject accuracy while emphasizing model interpretability [10].

These studies collectively demonstrate how machine learning techniques -ranging from traditional classifiers to advanced deep neural networks -enable efficient telemetry signal analysis, improving fatigue detection and potential real-time clinical monitoring (Table 1).

Table 1

Paper (Year)	**Core ML/DL Technique**	**Input Data Type**	**Key Metrics / Accuracy**	**Main Strengths / Innovation**	**Limitations Noted**
Kaveh et al. (2024)	SVM	Wireless Ear-EEG	93% Accuracy	Discreet, comfortable in-ear sensing; low motion artifacts	Small dataset; lab-based study
Zhu et al. (2021)	CNN (Inception-CNN)	EEG Spectrograms	95.6% Accuracy	Automatic feature extraction from 2D representations	Dataset size; potential overfitting
Zayed et al. (2025)	Hybrid CNN–SVM	EEG Scalograms	98.3% Accuracy	Combines strengths of DL feature learning and SVM classification	Computational cost of 2D conversion
Hassan et al. (2025)	Transformer (ViT)	Facial Video Frames	99.15% Accuracy	Contactless, robust to head movements; high accuracy	Privacy concerns; lighting dependency; high compute load

Chen et al. (2023)	Capsule Network	Multi-channel EEG	94.2% Accuracy	Better interpretability; captures spatial–temporal relations	Complex architecture; limited dataset
Zayed et al. (2024)	SVM + Feature Selection	EEG Data	99.8% Accuracy	Feature-optimized model achieving near-perfect detection	Limited subject diversity
Obaidan et al. (2024)	1D CNN (EEG_DMNet)	Raw EEG	97.03% Accuracy	Multi-scale feature extraction; efficient model	Dataset imbalance
Gao et al. (2023)	CNN–RNN	EEG Spectrograms	88.4% Accuracy	Sequential feature learning for fatigue detection	Moderate accuracy; needs optimization
Alghanim et al. (2024)	Hybrid Inception–ResNet	EEG Data	98.8% Accuracy	Highly generalized hybrid model	Model complexity; high training cost
Cui et al. (2022)	Compact CNN	Single-Channel EEG	73.2% Accuracy	Lightweight; interpretable	Cross-subject variability; reduced accuracy

4. Methodology of Review

The methodology involved systematic selection of research papers focusing on EEG-based ML telemetry systems between 2021 and 2025. Searches were conducted using IEEE Xplore, ScienceDirect, and PubMed with keywords such as *EEG telemetry, machine learning, drowsiness detection,* and *wearable sensors.* Each study was analyzed for model type, dataset used, preprocessing methods, and classification results. This qualitative synthesis identifies performance benchmarks and common algorithmic approaches used for telemetry signal interpretation.

5. Summary of Findings

The reviewed literature reveals several consistent patterns and achievements:

- **High Accuracy and Reliability:** Nearly all models achieved above 85% accuracy, with CNN-based models outperforming traditional ML algorithms.
- **Feature Extraction and Optimization:** Using spectrograms and wavelet scalograms improved performance by preserving both time and frequency features.
- **Reduced Electrode Dependency:** Some models demonstrated high accuracy using minimal electrodes (e.g., ear-EEG or single-channel setups).
- **Cross-Subject Adaptability:** Transfer learning and hybrid architectures improved generalization across individuals.
- **Energy Efficiency:** Lightweight models such as AGL-Net showed that ML algorithms could be optimized for embedded telemetry applications.

These findings confirm the practicality of integrating ML with biomedical telemetry for real-time monitoring and decision support.

6. Discussion of Limitations

While the reported accuracy rates are impressive, several challenges persist in the deployment of ML-based telemetry systems:

- **Limited Dataset Diversity:** Many studies use small, controlled datasets, which restrict model generalizability.
- **Inter-Subject Variability:** EEG signals vary widely across users, requiring calibration or adaptive algorithms.
- **Real-World Validation:** Most models are tested in laboratory environments rather than real-time clinical settings.
- **Hardware Constraints:** High computational requirements make it difficult to deploy complex DL models on portable telemetry devices.
- **Signal Quality Issues:** Movement artifacts and electrode noise can degrade signal clarity and classification accuracy.

Addressing these issues is vital for transitioning telemetry research from laboratory prototypes to clinical and consumer applications.

7. Future Work Directions

Future efforts should focus on the following areas:

- **Multimodal Data Fusion:** Integrating multiple biosignals (EEG, ECG, PPG, motion) can improve diagnostic accuracy.

- **Adaptive Learning Systems:** Developing self-calibrating ML models to handle individual variability.
- **Edge Computing and Energy Optimization:** Creating efficient AI models capable of running on wearable telemetry devices.
- **Data Sharing and Standardization:** Building large, annotated public datasets to support reproducibility and benchmarking.
- **Explainable AI:** Ensuring transparency in model decisions to enhance trust in clinical adoption.

Such advancements will bring telemetry closer to intelligent, patient-specific health monitoring systems.

8. Conclusion

This paper reviewed the current state of biomedical telemetry and highlighted how machine learning techniques are transforming remote patient monitoring. By analyzing ten recent EEG-based drowsiness detection studies, it was found that ML algorithms can classify physiological states with remarkable precision. Despite existing challenges like limited data and variability, progress in sensor miniaturization, algorithm optimization, and hardware integration promises a future where wearable telemetry systems deliver continuous, intelligent healthcare support. Machine learning has thus become a cornerstone of the next generation of biomedical telemetry—paving the way for accurate, accessible, and personalized patient monitoring.

References

1. I. A. Fouad, "A robust and efficient EEG-based drowsiness detection system using different machine learning algorithms," Ain Shams Engineering Journal, vol. 14, no. 3, p. 101895, 2023.
2. A. Zayed, N. Belhadj, K. Ben Khalifa, M. H. Bedoui, and C. Valderrama, "Efficient generalized EEG-based drowsiness detection approach with minimal electrodes," Sensors, vol. 24, no. 13, p. 4256, 2024.
3. A. Zayed, N. Belhadj, K. Ben Khalifa, C. A. V. Sakuyama, and M. H. Bedoui, "A novel hybrid approach for drowsiness detection using EEG scalograms to overcome inter-subject variability," Sensors, vol. 25, no. 17, p. 5530, 2025.
4. H. B. Obaidan, M. Hussain, and R. AlMajed, "EEG_DMNet: A deep multi-scale convolutional neural network for electroencephalography-based driver drowsiness detection," Electronics, vol. 13, no. 11, p. 2084, 2024.
5. D. Gao, X. Tang, M. Wan, G. Huang, and Y. Zhang, "EEG driving fatigue detection based on log-Mel spectrogram and convolutional recurrent neural networks," Frontiers in Neuroscience, vol. 17, p. 1136609, 2023.
6. M. Alghanim et al., "A hybrid deep neural network approach to recognize driving fatigue based on EEG signals," International Journal of Intelligent Systems, Article 9898333, 2024.
7. L. Guarda, J. E. Tapia, E. L. Droguett, and M. Ramos, "A Capsule Neural Network model for EEG-based drowsiness detection," Expert Systems with Applications, vol. 201, p. 116977, 2022.
8. H. Liu et al., "Attention-based multi-semantic graph convolutional network for EEG-based fatigue detection," Frontiers in Neuroscience, vol. 17, p. 1275065, 2023.
9. W. Fang, L. Tang, and J. Pan, "AGL-Net: An efficient neural network for EEG-based driver fatigue detection," Journal of Integrative Neuroscience, vol. 22, p. 146, 2023.
10. J. Cui, Z. Lan, Y. Liu, R. Li, and F. Li, "A compact and interpretable convolutional neural network for cross-subject driver drowsiness detection from single-channel EEG," Methods, vol. 202, pp. 173–184, 2022.

Exploring the Potential of Bacteria-Mediated Chromium Chelation in Biomedical Waste Contaminated Water

Kanthasami Vijaayganapathi, Saqib Hassan1, Kunal Biswas

Sathyabama Institute of Science and Technology, Tamilnadu, Chennai, India
kunalbiswas.irc@sathyabama.ac.in

Abstract. Chromium pollution in biomedical wastewater is a serious public and ecological health risk because of its high toxicity, carcinogenicity, and aquatic bioaccumulation potential. Traditional remediation processes like chemical precipitation, ion exchange, and membrane filtration are energy intensive, expensive, and produce secondary pollutants, thus restricting their use on a large scale and longevity. Bacteria-assisted chromium chelation presents a sustainable, green, and its effective bioremediation option for the toxic chromium species. Bacteria have inherent processes e.g., biosorption, bioaccumulation, enzymatic reduction, and secretion of extracellular polymeric substance (EPS) that allow chelation and immobilization of toxic chromium compounds. The cell wall of the bacteria, which is loaded with carboxyl, hydroxyl, amino, and phosphate moieties, favours binding of chromium ions and overall removal. Studies have established that bacterial species such as Pseudomonas, Bacillus, and Escherichia coli efficiently reduce the Chromium concentrations to environmental safe limits. In addition, incorporating bacterial bioremediation with nanomaterials, biofilm reactors, or genetic engineering greatly enhances the Chromium removal efficiency, stability, and adaptability under different wastewater conditions. This review highlights the mechanisms, benefits, and drawbacks of bacteria-mediated chromium chelation and its potential as a green, economical, and sustainable method for chromium-contaminated biomedical wastewater treatment and global environmental and the human health protection.

Keywords. Bacteria; Chromium; waste water; Chelation; remediation; Biomedical contaminated sites.

1. Introduction

Chromium pollution from industrial and biomedical effluents is a severe environmental and public health issue because of its toxic, mutagenic, and carcinogenic properties, especially in the hexavalent form (Cr^{6+}). Traditional techniques for chromium removal, including chemical precipitation, ion exchange, and membrane filtration, tend to be costly and produce secondary pollutants. Bacteria-mediated chelation of chromium provides a cost-effective and environmentally friendly solution for decontaminating water with chromium [1]. Some bacteria have the inherent capacity to convert poisonous Cr^{6+} to less toxic trivalent form (Cr^{3+}) or chelate chromium ions by extracellular polymeric substances and biosorption processes. Screening of the possibilities of these bacterial systems from biomedical polluted water sources may lead to the detection of effective strains for bioremediation under natural

conditions. This method not only supports sustainable wastewater treatment but also helps in the creation of green technologies for the restoration of the environment.

Biomedical technology, as applied to bacteria-mediated chromium chelation activities from biomedical contaminated water areas, is the utilization of biological and technological concepts to track, control, and abate pollution by biomedical waste. It entails the utilization of microbial systems and biotechnological instruments for detecting, isolating, and promoting bacteria with the ability to detoxify heavy metals such as chromium from biomedical effluents. A schematic of Graphical Abstract (GA) is shown in Fig 1 (prepared in biorender.com). This technology combines microbiology, environmental biotechnology, and biomedical waste management in the generation of environmentally friendly solutions for minimizing toxic metal levels. Through bacterial chelation and biotransformation processes, biomedical technology facilitates the secure treatment and recycling of polluted water, hence safeguarding ecosystems and public health [2]. Biomedical technology in the context of chromium chelation by bacteria involves the utilization of technological and biological principles to track, control, and clean up chromium contamination resulting from biomedical and healthcare waste streams. Biomedical wastewater generally comprises leftover chemicals, disinfectants, and heavy metals such as chromium that arise from laboratory operations, diagnostic reagents, medical device production, and hospital effluents [3]. Biomedical technology seeks to couple microbiological systems with contemporary engineering methods for the purpose of designing sustainable treatment strategies. On this occasion, naturally metal-binding and detoxifying bacteria are used as bio-agents to sequester and convert toxic chromium forms to less toxic analogues through an environmentally friendly alternative to traditional chemical treatment processes that tend to be costly and generate secondary pollutants [4]. In addition, bio-medi technology of bacteria-assisted chromium chelation involves the search and design development and optimization of biotechnological means like biofilm reactors, microbial consortia, and genetically engineered strains for achieving better chromium removal efficacy and flexibility in actual wastewater conditions. Utilizing microbial metabolism, enzyme systems, and extracellular polymeric substances, this technique enables simultaneous detoxification and recovery of precious metals from waste, thereby being consistent with circular bioeconomy principles. Accordingly, biomedical technology acts as an intermediary between microbiology, biotechnology, and environmental engineering—to facilitate the development of new, environmentally friendly, and economical solutions for chromium pollution alleviation from biomedical wastewater sources while protecting human and environmental health.[5].

2. Current Scope in bacteria-mediated Chromium chelation activities from biomedical contaminated water sites

The present scale of bacteria-mediated chromium chelation in biomedical-contaminated water ranges from laboratory mechanistic investigations to pilot-scale applications, with increasing focus on scaling up promising strains and consortia to integrated treatment units. In the laboratory bench, scientists are charting mechanisms (biosorption, enzymatic Cr(VI)→Cr(III) reduction, efflux, bioaccumulation and EPS-mediated sequestration) and designing strong species (e.g., Bacillus, Pseudomonas and mixed native consortia) which are resilient to common biomedical effluent stresses including high organic load, antibiotics and combined pollutants [6]. Applied work increasingly investigates biofilms, immobilized cells and mixed microbial consortia in order to enhance stability and reusability, and a number of investigations describe pilot-scale bioreactors or hybrid systems (biological + adsorption/chemical polishing) demonstrating good chromium removal under practical

conditions. Technological advances in omics, adaptive laboratory evolution and synthetic biology are applied to increase rates of reduction, tolerance to stress and range of substrates, but electron donors, redox mediators and co-treatments are also investigated in studies to increase in situ reduction activity. Even with these improvements, however, the technology still lags behind in scale-up issues (variable effluent composition, long-term performance, regulatory acceptance, and metal-laden biomass safe disposal), so the short-term scope is realistic: integrate bacterial chelation with standard polishing steps, focus pilot trials on actual biomedical wastewater, and apply genomics-led strain/consortium selection to close laboratory success and field deployment. The present range of bacteria-mediated chromium chelation at biomedical contaminated water locations is quickly widening as scientists investigate its viability as a cheap, non-polluting, and effective bioremediation method. Several bacterial species including Pseudomonas, Bacillus, Escherichia coli, and Enterobacter have been shown to be highly efficient in the removal of chromium by processes such as biosorption, bioaccumulation, and enzymatic reduction. These bacteria can transform lethal hexavalent chromium (Cr^{6+}) to its less harmful trivalent form (Cr^{3+}), thus minimizing environmental and health hazards posed by heavy metal contamination. Microbiological and biotechnological advancements have enabled the isolation, identification, and improvement of efficient chromium-resistant bacterial strains from contaminated environments, enhancing their adaptability and efficiency under different pH, temperature, and salinity conditions characteristic of biomedical wastewater [7]. Furthermore, combining bacterial chelation with newer technologies like nanomaterials, biofilm reactors, and genetic modification has broadened the horizon of this technique even more. Bacterial systems assisted with nanoparticles increase metal-binding capacity and electron transfer, whereas biofilm-based systems enhance microbial stability and continuous treatment efficacy [8]. Genetic alterations and synthetic biology resources are utilized to design bacterial strains with enhanced chromium reduction pathways, increased enzyme catalytic ability, and elevated metal tolerance. The emphasis now is also placed on the construction of scaled-up bioreactors and hybrid systems that integrate bacterial cleanup with physical or chemical treatments for efficient large-scale treatment of biomedical wastewater. In total, the widening reach of chromium chelation mediated by bacteria mirrors its increasing significance as a foundation of contemporary eco-biotechnological strategies in the quest for clean water, pollution abatement, and biomedically sound waste treatment [9].

3. Model of bacteria-mediated Chromium chelation activities from biomedical contaminated water sites

The bacteria-mediated chelation model of chromium from biomedical polluted water sources describes the sequential physicochemical and biological mechanisms of chromium detoxification by microbial systems. In this model, chromium-polluted effluents initially pass through bacterial interaction, wherein particular species of bacteria adsorb and chelate chromium ions via their cell surface functional groups like carboxyl, hydroxyl, and amino groups. This is succeeded by enzymatic reduction, wherein harmful hexavalent chromium (Cr (VI)) is reduced to the lower-toxicity trivalent form (Cr (III)), usually stabilized by chelation or formation of a complex with bacterial metabolites or extracellular polymeric substances. The process as a whole is controlled by environmental conditions such as pH, temperature, redox potential, and availability of organic matter. The model combines adsorption, biotransformation, and precipitation processes to outline the dynamic chromium removal from polluted water, hence predicting bacterial detoxification effectiveness under disparate environmental and operational conditions [10]. Bacteria-mediated chromium chelation model

from biomedical contaminated water sources is premised on the inherent metabolic and structural potential of bacteria to immobilize and detoxify substances like toxic chromium species, specifically hexavalent chromium (Cr^{6+}). Upon introduction of bacteria into chromium-polluted wastewater, their cell walls that contain abundant functional groups like carboxyl, hydroxyl, amino, and phosphate moieties enable the first adsorption and coordination of chromium ions through surface complexation and electrostatic forces. This biosorption process aids in quick dissolved chromium removal from the liquid phase. Subsequent to adsorption, some bacterial strains employ enzymatic reduction processes in which Cr^{6+} is reduced to less toxic trivalent Cr^{3+} by intracellular or membrane-bound reductase enzymes [11]. The lower chromium precipitates in the form of insoluble compounds or gets sequestered in the extracellular polymeric substances (EPS) released from bacteria, thus reducing its bioavailability and mobility in water systems. Along with these inherent processes, the model also includes synergistic interactions between environmental parameters and microbial populations that augment chromium elimination efficiency. Biofilm growth on solid supports enables bacteria to operate cooperatively in an EPS matrix produced by themselves, enhancing surface area for binding chromium and providing protection from hostile wastewater environments. The model may also be extended by incorporating other new technologies like nanomaterial supports or genetically engineered bacterial strains to further enhance chromium chelation capacity, enzymatic activity, and system stability. Generally speaking, this model displays a holistic and green bioremediation strategy that not only minimizes chromium toxicity in biomedical wastewater but also enhances sustainable heavy metal pollutant management through the synergistic effect of microbial metabolism, surface chemistry, and contemporary biotechnological advances.

4. Concept of Bacteria-Mediated Chromium Chelation

The mechanism of bacteria-mediated chromium chelation is the capacity of some microorganisms to detoxify, convert, and immobilize carcinogenic chromium species by undergoing biological and biochemical reactions. Chromium in the hexavalent state (Cr^{6+}) is extremely toxic, mutagenic, and carcinogenic and poses critical threats to aquatic ecological systems and human health when discharged through biomedical wastewater. Bacteria have natural defense and adaptation processes that allow them to survive and thrive in metal-polluted environments. Bacteria can adsorb and reduce chromium ions through mechanisms like biosorption, bioaccumulation, bioreduction, and extracellular polymeric substance (EPS) production [12]. In biosorption, chromium ions bind with the bacterial cell wall, which is composed of functional groups such as carboxyl, hydroxyl, phosphate, and amino moieties. At the same time, enzymatic reduction permits bacteria to reduce Cr^{6+} to the less toxic trivalent chromium (Cr^{3+}), precipitating or forming stable complexes and reducing its solubility and mobility in water systems. This bio-chelation process is both environmentally friendly and economical compared to other traditional physicochemical treatments. The effectiveness of bacterial chromium chelation relies on various parameters such as bacterial strain type, environmental conditions (ion concentration, temperature, and pH), and the availability of other organic or inorganic substances [13]. Some bacterial genera like Pseudomonas, Bacillus, Escherichia coli, and Enterobacter have been well researched for chromium-binding and reduction properties. In addition, biofilm development can reinforce the chelation process where bacterial populations synthesize an extracellular protective EPS matrix enhancing stability and stress resistance. In conclusion, the notion of bacteria-mediated chelation of chromium is a new biotechnological strategy for biomedical wastewater treatment that further supports pollution sustenance and the safeguarding of ecological and human health.

5. Biosorption and Bioaccumulation

Biosorption and bioaccumulation are the two major mechanisms through which bacteria remove and immobilize chromium from the contaminated environment. Biosorption is a non-active process involving the adsorption of chromium ions onto the bacterial cell surface, where the chromium ions attach to negatively charged functional groups on the cell wall [13]. This process does not involve metabolic energy and is primarily caused by electrostatic attraction, ion exchange, complexation, and surface precipitation. A typical schematic of the process of bioaccumulation and biosorption is shown in Fig. 1 (prepared in biorender.com). Bioaccumulation, in contrast, is an energetic process that involves the passage of chromium ions into the bacterial cell via membrane transport systems with subsequent intracellular sequestration or transformation. Bacteria for bioaccumulation are able to accumulate chromium within intracellular storage compartments, which decreases its bioavailability and toxicity in the external environment. These processes altogether contribute significantly to decreasing chromium concentration in biomedical wastewater.

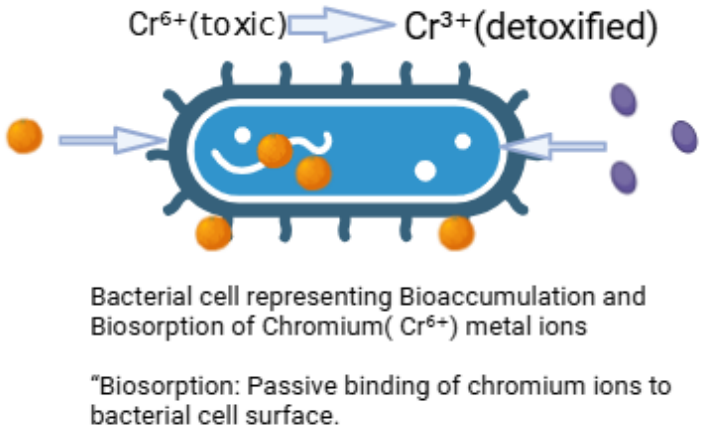


Fig. 1. Chromium Chelation by bacteria — Biosorption and Bioaccumulation mechanism

6. Enzymatic Reduction of Cr^{6+} to Cr^{3+}

Enzymatic reduction of hexavalent chromium (Cr^{6+}) to trivalent chromium (Cr^{3+}) is an important detoxification process in bacterial chromium chelation [14]. This decrease is facilitated by chromate reductases in bacteria's cytoplasm and periplasmic regions. Cr^{6+} is an electron acceptor used in bacterial metabolism, with reducing substances like NADH or NADPH donating electrons to reduce Cr^{6+} to the less toxic, less soluble form of Cr^{3+}. A schematic of the process is depicted in Fig. 2 (prepared in biorender.com). The formed Cr^{3+} is prone to precipitate or stabilize complexes with organic and inorganic molecules, facilitating immobilization. Through this enzymatic action, chromium is not only detoxified but also

enables bacteria to resist metal stress, ensuring their survival in contaminated biomedical wastewater.

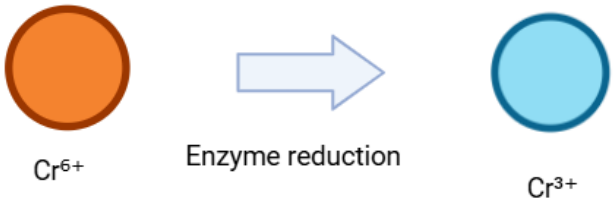


Fig 2. Enzymatic reduction process of **Cr^{6+}** to **Cr^{3+}**

7. Role of Extracellular Polymeric Substances (EPS)

Bacteria-secreted Extracellular Polymeric Substances (EPS) are crucial in chelating chromium by functioning as a natural bio-sorbent and defensive barrier. EPS are made up of polysaccharides, proteins, lipids, and nucleic acids with more than one functional group available for binding metal ions [12]. A typical schematic of the role of EPS is shown in Fig.3 (prepared in biorender.com). The EPS matrix creates a hydrated, sticky surface that promotes the adhesion of bacteria to surfaces and allows for the entrapment of chromium ions. EPS also shields bacterial cells from oxidative stress due to Cr^{6+} toxicity, aids in biofilm formation, and enhances the general metal tolerance of bacterial populations. The presence of EPS greatly enhances available surface area for chromium adsorption and hence the removal efficiency in wastewater systems.

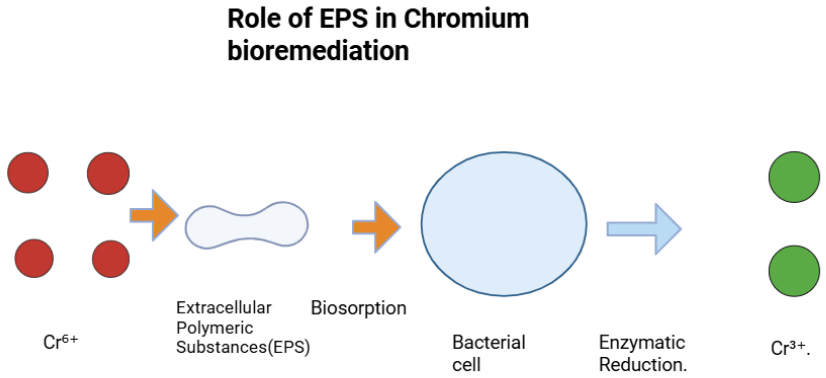


Fig. 3. Implication of EPS in chromium ion chelation cum reduction in bacterial mediated process

8. Functional Groups Responsible for Binding of Metals

The EPS matrix and bacterial cell wall both have various functional groups responsible for the chelation and binding of chromium ions. These functional groups include carboxyl (-COOH), hydroxyl (-OH), amino (-NH_2), phosphate (-PO_4^{3-}), and sulfhydryl (-SH) groups [12]. A typical schematic demonstrating the mechanism of chromium ions absorption to functional groups at EPS surface in bacteria mediated chromium chelation is shown in Fig. 4 (prepare in biorender.com). These functional moieties present active sites for the interaction of chromium ions through coordination, ion exchange, and covalent binding. For example, carboxyl and phosphate groups tend to coordinate positively charged chromium species by electrostatic forces, while amino and hydroxyl groups coordinate through coordination complexes. The number and distribution of these functional groups on bacterial surfaces dictate the metal-binding capacity and the overall efficiency of chromium removal. Collectively, these functional groups make bacterial systems highly effective bio-chelators for treating chromium-contaminated biomedical wastewater.

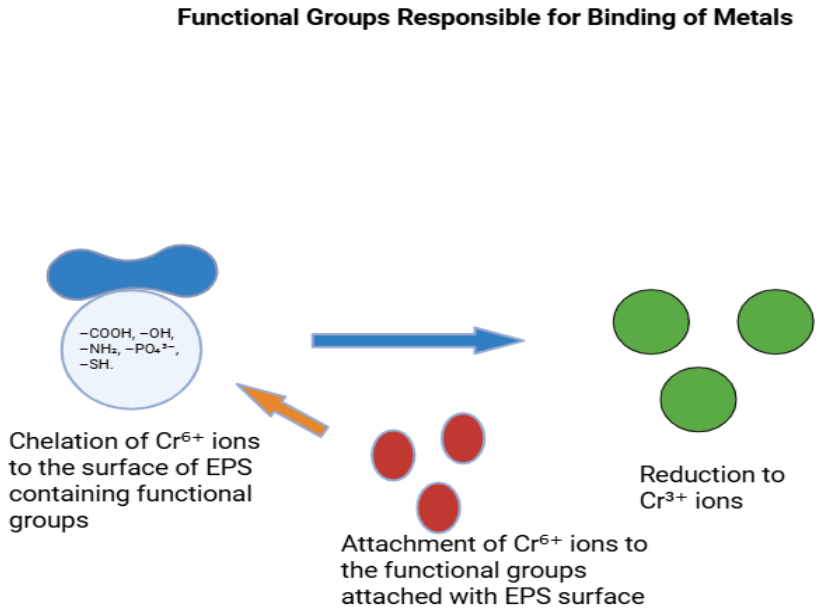


Fig. 4. Role of Functional groups in Chromium chelation mechanism using bacteria

9. Bacterial Strains with Chelation Potential

Some bacterial strains have very high chromium chelation capacity because they can tolerate, bind, and reduce toxic chromium species to less toxic ones. Some of the most researched include *Pseudomonas, Bacillus*, *Escherichia coli, Enterobacter, and Acinetobacter species*. They are highly adaptable to polluted environments and have varying metabolic capabilities that allow for chromium removal via biosorption, bioaccumulation, and enzymatic reduction [**4, 7**]. For instance, *Pseudomonas aeruginosa and Bacillus subtilis* have been reported to convert hexavalent chromium (Cr^{6+}) into trivalent chromium (Cr^{3+}) with high

efficiency due to the availability of chromate reductase enzymes and cell wall structures that are strong and rich in active binding sites. Equally, *Escherichia coli* utilizes intracellular redox processes for detoxification of Cr^{6+}, whereas *Enterobacter cloacae* is characterized by high tolerance and chromium-binding capacity via its extracellular polymeric substances (EPS) that facilitate improved metal ion sequestration and cell protection against toxicity. Existing research has also cantered on the possibility of mixed microbial consortia and genetically modified bacterial strains for efficient chromium chelation of biomedical wastewater. The cooperation among various species of bacteria in consortia enables enhanced chromium removal due to mutualistic metabolic pathways and enhanced biofilm strength. Genetic engineering of bacteria like *Pseudomonas putida and Bacillus megaterium* has been investigated to overexpress reductase enzymes of chromium or enhance the production of EPS so that metal-binding efficiency and stress resilience can be enhanced. In addition, advancements in nanobiotechnology and immobilized bacterial systems have further improved the practical usability and resilience of these strains to changing wastewater conditions [7]. In general, these bacterial strains are a promising, environmentally friendly solution for effective chromium-contaminated biomedical wastewater treatment in favour of sustainable and biologically mediated remediation practices.

10. Biomedical Technology Applications

Biomedical technology uses of bacteria-mediated chromium chelation are centered around combining microbial systems with engineered and biotechnological devices to enhance the reliability and efficiency of chromium removal from biomedical wastewater [15]. The majority of applications for immobilized bacterial systems and biofilm reactors are based on supporting microbial community growth in a controlled environment, enabling bacteria to settle on surfaces, develop protective biofilms, and achieve high cell density for repeated wastewater treatment. These systems improve chromium adsorption and biotransformation by presenting a stable microenvironment that shields bacteria from the stress of toxic metals and allows for extended activity. Furthermore, utilization of sophisticated bioreactors allows for optimization of conditions like pH, temperature, and availability of nutrients, further enhancing the removal efficiency of toxic chromium species [5]. Other emerging biomedical technology strategies involve the application of nanomaterials and genetic manipulation to enhance bacterial chelation capacity. Nanoparticles, when mixed with bacterial cells, enhance electron transfer, redox reactions, and adsorption capacity for chromium ions. Genetic and metabolic engineering strategies enable the creation of bacterial strains with increased chromate reductase activity, heavy metal tolerance, and the production of extracellular polymeric substances (EPS), which enable metal binding and immobilization. These integrated techno-strategies render bacterial chromium chelation more scalable, stable, and responsive to varying wastewater conditions, offering an environmentally friendly, cost-effective, and sustainable approach for biomedical wastewater treatment and prevention of environmental and human health hazards of heavy metal pollution.

11. Current Scope and Research Advances

The existing range of chromium chelation mediated by bacteria has extensively expanded with scientists persistently investigating microbial capability in sustainable biomedical wastewater treatment. Experiments in recent past have shown that several bacterial genera, such as Pseudomonas, Bacillus, Escherichia coli, Enterobacter, and Acinetobacter, exhibit satisfactory chromium tolerance and removal capacities via biosorption, bioaccumulation, and enzymatic reduction. These microorganisms can thrive under stressful

wastewater environments and efficiently convert toxic hexavalent chromium (Cr^{6+}) to its less toxic trivalent species (Cr^{3+}) to reduce its environmental impact. The growing knowledge of bacterial metabolism, enzyme kinetics, and cell wall chemistry has also enabled enhanced selection and optimization of microbial cultures for particular wastewater environments. In addition, mixed bacterial consortia have come into the limelight, as mixed microbial synergies tend to increase chromium detoxification performance and stability over single-strain systems. Recent developments also focus on integration of bacterial remediation with innovative technologies to address limitations including poor performance under conditions of extreme pH or temperature and sluggish reduction rates. The integration of nanomaterials into bacterial systems has been effective in promoting chromium adsorption and redox reactions through optimized surface interaction and electron transfer. Likewise, biofilm reactors and immobilized bacterial systems provide higher resistance against toxic metal concentrations and allow continuous operation treatment. Genetic engineering and synthetic biology are further transforming this discipline by creating altered bacterial strains with overexpressed chromate reductase enzymes, increased EPS production, and enhanced resistance to toxic conditions. Generally, these developments suggest an increasing capability for bacteria-assisted chromium chelation to emerge as a mass-level, eco-friendly means of biomedical wastewater treatment towards global sustainability and environmental conservation efforts.

12. Challenges and Limitations

Not withstanding its huge potential, bacteria-mediated chromium chelation suffers from a number of challenges and limitations that are restraining its large-scale application to biomedical wastewater treatment. One of the biggest challenges is the sensitivity of the bacterial system to environmental factors like pH, temperature, and the availability of competing ions or toxic compounds in wastewater. Changes in these variables have a significant impact on the growth of bacteria, enzyme activity, and efficiency of chromium reduction. In addition, it is challenging to retain the viability and metabolic activity of bacteria in the presence of high levels of chromium or co-contaminating heavy metal pollutants because high toxicity will harm cellular components and suppress chelating activities. The slow kinetics of biological reduction is another limitation compared to chemical treatments, which can slow down the overall treatment rate in high-capacity wastewater systems. In addition, upscaling from laboratory to large-scale industrial application continues to be problematic owing to operational issues and the demands of maintaining controlled conditions in order to support bacterial performance. A further major limitation is concerned with long-term stability and preservation of bacterial cultures within bioreactors or natural wastewater treatment systems. Ongoing exposure to changing waste compositions and stress conditions may result in reduced microbial activity or metal-removal efficiency loss over time. The challenges associated with the safe disposal and handling of the treated metal-loaded biomass are also present, as unsafe disposal may result in secondary pollution. Maintaining bacterial consortia, optimizing nutrient supply, and ensuring performance robustness under diverse wastewater conditions are also issues of economic viability. In order to overcome these challenges, recent studies are aiming to increase bacterial stability and tolerance using genetic modification, immobilization procedures, and incorporation of hybrid systems that utilize biological and physicochemical processes. Overcoming these constraints will be instrumental in developing bacteria-based chromium chelation as a dependable, scalable, and environmentally friendly technology for biomedical wastewater treatment.

13. Future Prospects and Research Directions

The prospects for future applications of bacteria-mediated chelation of chromium are very promising, with current research further developing its efficiency, flexibility, and scalability in practical wastewater treatment. A major direction of current research is the genetic and metabolic modification of bacterial strains to enhance their tolerance of chromium, enzymatic reducing ability, and the production of extracellular polymeric substance (EPS). By controlling certain genes involved in the uptake and reduction of chromium, researchers can engineer bacteria with improved detoxification and survival at harsh wastewater conditions. Moreover, breakthroughs in synthetic biology and omics sciences (genomics, proteomics, and metabolomics) are assisting in deciphering microbial pathways and refining bacterial metabolism for better chelation. Another new frontier is microbial consortia and biofilm-based systems, which facilitate cooperative interactions among various bacterial species to achieve greater stability, robustness, and metal-removal efficacy.

Future work also seeks to combine bacterial chelation with nanotechnology, membrane systems, and bio-electrochemical reactors to develop hybrid bioremediation platforms with improved chromium removal performance. Metal adsorption, enzyme activity, and redox reactions can be enhanced by nanoparticle-assisted bacterial systems, whereas bio-electrochemical systems enable the generation of energy from treatment, thereby increasing the sustainability. Pilot-scale bioreactor designs and in-situ bioremediation approaches in biomedical wastewater treatment plants will be studied further to take laboratory successes to practical levels. Additionally, establishing eco-safe strategies for metal recovery and biomass reuse might provide economic benefits and favour circular bioeconomy models. In summary, the prospects for bacteria-mediated chromium chelation are multidisciplinary innovation—the synergy of microbiology, biotechnology, and environmental engineering—to realize effective, clean, and affordable methods for preventing heavy metal contamination and ensuring environmental and human well-being.

14. Conclusion

Bacteria-assisted chromium chelation is an effective, green, and sustainable method for the elimination of toxic chromium species from biomedical contaminated wastewater. Different bacterial strains like Pseudomonas, Bacillus, Escherichia coli, and Enterobacter showed high chromium-removing capability through mechanisms such as biosorption, bioaccumulation, enzymatic reduction, and the secretion of extracellular polymeric substances (EPS). These biological processes efficiently reduce hexavalent chromium (Cr^{6+}) to its less toxic trivalent state (Cr^{3+}), thus lowering its solubility and bioavailability. Incorporation of sophisticated technologies like nanomaterials, biofilm reactors, and genetic engineering has also improved bacterial chelation efficiency, responsiveness, and system stability under fluctuating wastewater conditions. Though there are challenges in scaling up the process and preserving bacterial activity in extreme conditions, recent studies present exciting developments that make bacterial chelation a potential alternative to traditional physicochemical processes. The effective use of bacteria-mediated chromium chelation carries important implications for environmental and human health. By minimizing chromium toxic concentrations in biomedical wastewater, the process inhibits aquatic ecosystem contamination, safeguards biodiversity, and reduces heavy metal entry into the food chain. Detoxification of Cr^{6+} also reduces the threats of carcinogenic, mutagenic, and organ-disrupting effects on humans using contaminated water sources. In addition, as an eco-friendly and low-cost technology, bacterial chelation follows the international sustainability goals and environmental protection regulations to minimize industrial contamination. Overall, using

bacteria-mediated chromium chelation within wastewater treatment processes is conducive to cleaner water resources, ecological balance, and protecting public health and the environment in the long term.

Acknowledgement

Authors are thankful to the Centre for Nanoscience and Nanotechnology, International Research Centre and Department of Biotechnology of Sathyabama Institute of Science and Technology, Chennai, India for providing support in completion of the study.

References

1. Kanmani, P., Aravind, J. & Preston, D., 2012. Remediation of chromium contaminants using bacteria. International Journal of Environmental Science and Technology, 9, pp.183–193.
2. Khan, M.S.A., 2024. Applications of bioremediation in biomedical waste management: current and future prospects. Brazilian Archives of Biology and Technology, 67, p.e24230161.
3. Jindal, A.K., Gupta, A., Grewal, V.S. and Mahen, A., 2013. Biomedical waste disposal: A systems analysis. medical journal armed forces india, 69(4), pp.351-356.
4. Elahi, A., Arooj, I., Bukhari, D.A. and Rehman, A., 2020. Successive use of microorganisms to remove chromium from wastewater. Applied Microbiology and Biotechnology, 104(9), pp.3729-3743.
5. Khattar, J.I., Sarma, T.A. and Sharma, A., 2007. Optimization of chromium removal by the chromium resistant mutant of the cyanobacterium Anacystis nidulans in a continuous flow bioreactor. Journal of Chemical Technology & Biotechnology: International Research in Process, Environmental & Clean Technology, 82(7), pp.652-657.
6. Srinath, T., Verma, T., Ramteke, P.W. and Garg, S.K., 2002. Chromium (VI) biosorption and bioaccumulation by chromate resistant bacteria. Chemosphere, 48(4), pp.427-435.
7. Ozturk, S., Kaya, T., Aslim, B. and Tan, S., 2012. Removal and reduction of chromium by Pseudomonas spp. and their correlation to rhamnolipid production. Journal of hazardous materials, 231, pp.64-69.
8. Kumar, H., Sinha, S.K., Goud, V.V. and Das, S., 2019. Removal of Cr (VI) by magnetic iron oxide nanoparticles synthesized from extracellular polymeric substances of chromium resistant acid-tolerant bacterium Lysinibacillus sphaericus RTA-01. Journal of Environmental Health Science and Engineering, 17(2), pp.1001-1016.
9. Bose, R., Kulkarni, S., Lakkakula, J. and Wagh, N.S., 2023. Genetic Engineering for Chromium Removal. In Chromium in Plants and Environment (pp. 139-166). Cham: Springer Nature Switzerland.
10. Narayani, M. and Shetty, K.V., 2013. Chromium-resistant bacteria and their environmental condition for hexavalent chromium removal: a review. Critical Reviews in Environmental Science and Technology, 43(9), pp.955-1009.
11. Chen, J. and Tian, Y., 2021. Hexavalent chromium reducing bacteria: mechanism of reduction and characteristics. Environmental Science and Pollution Research, 28(17), pp.20981-20997.
12. Batool, R., Marghoob, U. and Kalsoom, A., 2017. Estimation of exopolysaccharides (EPS) producing ability of Cr (VI) resistant bacterial strains from tannery effluent. Journal of Basic & Applied Sciences, 13, pp.589-596.
13. Alam, M.Z. and Ahmad, S., 2011. Chromium removal through biosorption and bioaccumulation by bacteria from tannery effluents contaminated soil. Clean–Soil, Air, Water, 39(3), pp.226-237.
14. Garg, A., Chauhan, P., Kaur, C., Perveen, S., Arora, P.K., Garg, S.K., Singh, V.P. and Srivastava, A., 2025. Bacterial bioremediation strategies for heavy metal detoxification: a multidisciplinary approach. Environmental Sustainability, pp.1-23.
15. Alum, E.U. and Uti, D.E., 2025. Modern perspectives on chelation therapy: optimizing biochemical approaches to heavy metal detoxification. Toxicology and Environmental Health Sciences, pp.1-14.

The Digital Transformation of Talent Acquisition: A Review of AI-Driven Strategies and Ethical Considerations

Gokul Nittin A, Sivakumar Rajagopal

Vellore Institute of Technology, Vellore, Tamil Nadu, India
gokulnittin.a2022@vit.ac.in.

Abstract. The paper reviews current trends in talent acquisition, focusing on the digital transformation from traditional to modern recruitment and selection processes. It traces the emergence and development of technology-driven strategies, including e-recruitment, social media sourcing, gamification, and the widespread integration of Artificial Intelligence (AI). The pursuit of operational efficiency through automation is examined in conjunction with the critical need for fairness, bias mitigation, and an enhanced candidate experience. Basic recruitment trends are tracked, including the use of AI for automated resume screening, predictive analytics for candidate matching, and chatbots for applicant engagement. The strategic value of structured interviews, employer branding, and inclusive hiring practices are also considered. A conceptual framework for evaluating the integration of these technologies is implicitly proposed, centered on balancing technological capability with ethical governance. For this purpose, the tension between data-driven objectivity and the risk of algorithmic bias is analyzed. Promising directions for future research are outlined, emphasizing the need for robust auditing of AI systems, exploring hybrid models that retain the human element, and examining the long-term impact of digital tools on workplace diversity and organizational culture.

Keywords. Digital recruitment, artificial intelligence, talent acquisition, algorithmic bias, ethical HRM.

1. Introduction

Attracting, assessing, and acquiring talent is an organizational success pillar and a major factor of competitive advantage in the contemporary, dynamic digital economy. Within the past twenty years, talent acquisition has been shifting towards a paradigm change, leaving behind the old, mostly manual and reactive approaches to talent acquisition and transitioning towards a digitally focused approach. This transformation is defined by the adoption of advanced digital solutions and, most importantly, the use of Artificial Intelligence (AI) to automate and streamline all the phases of the recruitment process and transform Human Resource Management (HRM) not merely into an administrative task but a business value generator.

This paper aims at revising the most important trends in this digital transformation. We combine the results of the latest literature to examine the contemporary methods of recruitment and selection, with particular attention to the strengths, advantages, and limitations of the AI-based strategies. It is the long-term aim to present a concise picture of the state of the current environment, discussing the formation of talent acquisition by technologies such as predictive analytics, chatbots, and automated screening and raising the issue of critical ethical issues, including algorithmic bias and data privacy, to the surface. In this review, the transition of traditional techniques to strategic processes, technology-enabled processes, the effects of the main digital tools on the efficiency and applicant experience, and the necessity of balancing

the idea of technological progress with fairness will be traced and the need to achieve an equal and effective recruitment ecosystem will be discussed.

2. The Evolution from Traditional to Strategic Recruitment

Recruitment in the past has been mostly considered a reactionary and operational process with the sole objective of filling vacancies that are arising at that moment. The manual sifting of resumes, a high dependency on unstructured interviews that are very vulnerable to the human cognitive biases, and a short-term perspective of staffing demand were the main characteristics of this traditional paradigm. The job advertisements were generic, and the process of screening was not very consistent, and the decision made was largely based on the intuition instead of the numbers.

Nevertheless, in the current hyper competitive global talent market, organizations are gradually acknowledging recruitment as not only a support service but as an important strategic business necessity which needs to be closely linked to long-term business goals. Strategic recruitment is, thus, characterized by proactive, data-driven and future-oriented nature and is concerned about the development and sustenance of a sustainable talent pipeline to address both the current and future needs of the organization. The planned development of an employer brand and the methodical exploitation of employee referrals are two main pillars that are an indication of this strategic change.

Employer Branding This entails proactively creating and selling a powerful, positive image of an employer of choice to attract, engage and keep the best talent. An attractive employer brand is an effective way of conveying the culture, values and employee value proposition of the organization to the potential and existing employees. The studies have shown that a good employer brand could positively affect employee satisfaction, engagement, turnover intention, and organizational citizenship behavior. It is an important distinguishing factor which helps to attract candidates who not only fit the skill requirements, but also their mission and ethos.

One of the most effective methods of recruitment has always been Employee Referrals, which involves taking advantage of social networks of existing employees to help bring potential employees. Referred employees tend to perform better at work, have an increased tenure period, record improved job attitude and tend to assimilate more easily into the organizational culture than the non-referred employees. The success of this can be attributed to several reasons: existing employees are able to give realistic job previews, are able to pre-screen prospective employees about their abilities and appropriateness in the culture and provide a signal of trust to the candidate and the company. In most organizations, employee referrals have a substantial proportion of employees who are hired thus their practicality. This shift in strategic thinking focuses on long-term planning, brand development, network leveraging, and data extraction offers the fundamental basis of the need to adopt and efficiently use the updated digital instruments that currently define the modern talent acquisition process. Such mechanisms can help companies to implement strategic projects such as employer branding and referral programs massively and more efficiently than ever before.

3. Key Digital Tools in Modern Talent Acquisition

A wide range of technological advances has helped to speed up and fundamentally change the shift to strategic, digitally enabled recruitment accelerated by substituting traditional processes, often paper-based ones. The COVID-19 pandemic was one of the events that accelerated the use of virtual interviews and remote assessments to an extent that

guaranteed the use of technology in modern hiring and resulted in the popularity of hybrid recruitment models.

3.1. E-Recruitment, Social Media, and Gamification

The first digital transformation saw the shift of recruitment to the online environment. Online job recruitment and company career sites had been the norm in advertising job openings and in receiving applications in electronic form. At the same time, the use of professional social media networks specifically LinkedIn transformed the process of sourcing candidates and now allows recruiters to access large pools of active and passive talent, allowing them to reach out to these individuals and network with them.

Later in the day, gamification has also come in as a new method that aims at increasing engagement and evaluation of the candidates. The game mechanics, challenges, and simulations could be added to the application or testing process so that the companies could make the experience more active, interactive, and enjoyable to the candidates. Through gamified tests, it is possible to get some insight into the problem-solving skills and creativity of candidates, time management, and behavioral characteristics through observing their performances in simulated on-the-job situations. This strategy can possibly provide a more objective estimate than traditional ones and enhance the brand image of the employer, especially to the younger demographics.

3.2. The Role of Artificial Intelligence (AI)

AI is the next frontier of the digitalization of talent acquisition that provides advanced features that automate, optimize, and offer intelligence to the whole recruitment funnel. Sourcing and Screening Resumes Resume screening AI algorithms are very useful in automating the time-consuming procedure of screening resumes. AI tools can read through thousands of applications, identify the relevant skills and experiences and then align them with the job requirement with contextual awareness than simple keyword searches using a combination of Natural Language Processing (NLP) and machine learning and rank the most qualified applicants. This brings high performance, saves vast amounts of time to hire measures, and liberates human recruiters to engage in more strategic and value-creating activities such as candidate relationship building. Candidate Engagement A chatbot is an AI-driven solution that is becoming more popular with companies to improve the candidate's experience by creating instant, 24/7 support. Such virtual assistants can respond to common questions of the candidate, give application status updates, prescreen basic qualification, and even help in setting up the interview. This guarantees a timely, consistent, and sometimes personalized communication that is significant in sustaining the interest of the candidates and creating an image of a modern employer. Assessment and Interviews such as While research shows that recruiters in general are not keen on eliminating human interaction in the interview process, AI is being used in assessment. Online tests and assessments are administered by use of technology and more AI tools, although the validity and ethics of such analysis are contested, are being developed to analyze video interviews in terms of communication patterns, sentiment, and non-verbal cues. Interestingly, it can be proposed that some candidates can feel less performance anxiety and be more at ease during the first stages of the selection process with AI-assisted or chatbot-facilitated interviews than in face- to-face environments. Predictive Analytics predictions such as by using large volumes of historical hiring data (e.g., candidate profiles, assessment results) and post-hiring performance data (e.g. performance reviews, tenure, promotions) to infer patterns and correlations thereby helping predict whether a particular candidate will be successful or retained in a job or within the organizational culture. This allows a more data-driven, strategic hiring process which enhances the quality of hire overall and minimizes expensive turnover.

4. Ethical Considerations and Bias Mitigation

Though the introduction of the digital tools, especially AI, has a massive potential to improve the efficiency, coverage, and anticipatory abilities of the talent acquisition process, it also comes with a lot of ethical issues that should be addressed with critical scrutiny and actively handled. The main challenge is to strike the right balance between the advantages of automation and data analysis and the principle of fairness, transparency, and dignity towards individuals.

4.1. The Duality of Bias

One of the challenges is the recruitment bias. Conventional recruiting techniques particularly unstructured interviews and subjective resume scanning are infamously open to a slew of human mental biases. These are affinity bias (giving preference to people who are close to us), the halo/horn effects (allowing a single positive/negative attribute to dominate the overall judgment) and confirmatory bias (focusing on information that supports our beliefs), and unconscious stereotypical inclinations based around safeguarding attributes such as gender, race, age, and disability.

On paper, AI provides a possible remedy to such human biases. Algorithms can be designed to disregard demographic data and concentrate on objective, job related factors, skills, experience and assessment scores, thereby encouraging merit-based assessment.

The danger is algorithmic bias. The AI systems are trained on the data they learn. In case such historical data signalizes previous discriminatory hiring practices or social injustices (e.g., lack of representation of some population in particular positions), the AI can become trained, institutionalized and even enhanced by said biases on a larger scale. A widely publicized example of an experimental recruiting tool used by Amazon, which was forced to be abandoned after it became trained to rank off resumes with the word women's and prefer men because it had been trained on the historically male-dominated data of applicants, is a bright example of a warning sign.

4.2. Strategies for Promoting Fairness

The complicated method of the issue is necessary to create the real fair and balanced hiring process and address the potential bias of the human decision and technological processes. The Structured Interviews, Inclusive Hiring Practices and Ethical AI Governance are the main strategies that have been identified in the literature. Structured Interviews It involves using standardized and job-related questions that are always directed to applicants, and a preset, behaviorally anchored scoring rubric, is one of the strongest and most empirically validated entry modes of enhancing objectivity, reliability, and predictive validity of interviews and minimizing the impact of interviewer bias. A Beyond designed interview, a group of practices should be implemented by organizations with the goal of expanding the diversity and limiting prejudice at all stages of the funnel. These are listed as blind resume screening (anonymizing applications by removing names, gender indicators and other demographic information and then reviewing them) and forming diverse hiring panels so that a variety of views can be factored into the assessment of the candidate, and targeted outreach to actively recruit members of underrepresented groups and atypical talent pools. It is also important to create non-discriminatory job descriptions in a neutral manner.

Ethical AI Governance such as Fair: In AI-based recruitment, requires a robust governance system that is founded on transparency, accountability, and oversight. This involves the periodic auditing of algorithms to actively identify and prevent possible discriminations in favor of the guarded categories. Giving the candidates an understanding of how AI is utilized during the assessment. Making sure that no critical hiring decisions are

achieved using algorithms only, a critical human-in-the-loop should be provided to review, contextualize, and make final decisions.

4.3. Data Privacy

The large amount of data gathered and processed about the applicants of digital recruitment platforms poses a major ethical and legal burden of concerns on the issue of data privacy. Companies should guarantee high adherence to the changing data protection laws, including European Union General Data Protection Regulation (GDPR). These rules will require the open data collection process, the safe storage, restrictions on the use of data, and give the applicants certain rights to their personal data, such as access, correction, and erasure. Fostering trust in the candidates depends on the ability to show them that there is a desire to be responsible and ethical in handling their sensitive data.

5. A Framework for Modern Recruitment

The evidence that is designed and assembled based on the analyzed literature tends to indicate that the most efficient and ethical paradigm of talent acquisition in the contemporary context is not an all-inclusive takeover by technology, but a mutually reinforcing hybridization model. The strategy is a focus on the effective use of technology to achieve more efficiency, objective, and reach, with a conscious attempt to retain and treasure human judgment in activities that need subtlety, empathy, and sophisticated relationship evaluation.

At the higher end of the recruitment funnel with massive data, repetitive tasks, and high volume, AI and automation prove to be the most valuable. This incorporates Sourcing applicants through large online databanks and business contacts. First-level resume screening and application screening based on objective, pre-set conditions. Responding to regular questions by the candidates and arranging appointments through chatbots.

Technology saves time which would otherwise be used by human recruiters by automating all these stages. This enables them to divert their attention to later phases of the funnel, which involves human capabilities in different aspects (high touch). These involve developing meaningful relations with prospective employees. Interviewing candidates based on their complex behavioral competencies and motivations through in-depth and structured behavioral interviews.

Assessment of the more important but less tangible factors of cultural fitness and compatibility with team dynamics. Coming to the final and multifaceted and context-sensitive decision on hiring, combining information understanding and professional opinion. It is not only necessary to create well-rounded hiring decisions using this human-centered approach, especially during the final selection steps, but also critical in providing a positive candidate experience, which most recruitment experts overwhelmingly favor in the recent studies.

The feasibility and effectiveness of such hybrid systems, which was required and improved in the circumstances of the COVID-19 pandemic, was proved by its practical implementation.

In the end, this framework proposal aims at achieving a perfect equilibrium. It takes advantage of the irrefutable efficiency benefits, data crunching capacity, and supposed bias minimization that technology promises, and protects the invaluable quality of human interaction, empathy, moral supervision and subtle judgment in the vital task of picking individuals to organizations.

6. Conclusion

According to the review, there is a clear indication of the possibility to map fields to search specific quantitative models of sustainability. Such models must be founded on thorough qualitative research of the concepts encompassing the contemporary meaning of corporate sustainability. One of such models is the one discussed above. The approach to stability analyses which is based on a set of ordinary differential equations offers numerous strengths which include the required information can be easily accessible within the organization Upping to modern IT the information can be easily processed It is not needed to solve the system to come up with the most appropriate values of the variables. To this end, an adequate method of defining the boundaries of system stability is the topological methods. It is on these grounds that the given model is a promising field in future research.

References

1. Schlachter S.D., J.R. Pieper, Employee Referral Hiring in Organizations: An Integrative Conceptual Review, Model, and Agenda for Future Research, Journal of Applied Psychology, 104, 11, 1325-1346, 2019.
2. Hegde A.S., S.C. Dwibedy, A Comparative Analysis of Traditional vs AI Drive Recruitment Methods, DHARANA - International Journal of Business, 15, 1, 3-10, 2024.
3. Vozar V., N. Jankelova, N. Jankelova, S. Moricova, Trends in the Recruitment and Hiring of Employees in the Healthcare Sector, Clinical Social Work and Health Intervention, 14, 4, 61-67, 2023.
4. Sankar S., K.R.M. Babu, D.S. Patel, M.M. Muthu, L. Arunachalam, A Study On Recruitment And Selection Process For Hiring The Potential Employees In Apollo Hospitals, Ilkogretim Online - Elementary Education Online, 19, 2, 1869-1881, 2020.
5. Singh R.P., Employer Branding Literature Review, Feedforward: Journal of Human Resource, 1, 2, 105-112, 2021.
6. Magdalenić D., L. Luić, Assessing the Impact of Digital Tools on the Recruitment Process Using the Design Thinking Methodology, Administrative Sciences, 15, 4, 139, 2025.
7. Dept of Psych EDI committee, McGill University, Structured vs. Unstructured Interview: Improving Accuracy & Objectivity, McGill University, Montreal, n.d.
8. Nascimento R.G.D.C.B.D., G.B. Marcon, J.M.F. Antunes Neto, The Strategic Recruitment Process And Its Impacts On Achieving Organizational Objectives: Bibliographic Review, IOSR Journal of Business and Management (IOSR-JBM), 26, 10 (Ser. 16), 01-14, 2024.
9. Kirovska Z., S. Josimovski, M. Kiselicki, Modern trends of recruitment - introducing the concept of gamification, Journal of Sustainable Development, 10, 24, 55-65, 2020.
10. Potočnik K., N.R. Anderson, M. Born, M. Kleinmann, I. Nikolaou, Paving the way for research in recruitment and selection: recent developments, challenges and future opportunities, European Journal of Work and Organizational Psychology, 30, 2, 159-174, 2021.
11. Tuttle L., K. Critchlow, Digital transformation in talent acquisition: Modern approaches to recruitment and selection, Int. J. of Research in Human Resource Management, 7, 1, 351-357, 2025
12. Edo-Osagie E.O., J.O. Ejechi, BRIDGING DIVERSITY: INCLUSIVE HIRING STRATEGIES AND ORGANISATIONAL STRUCTURE, UNIBEN Journal of Human Resource Management, 3, 1, 60-74, 2024.
13. Amaliyah, F.R.D. Cahyo, D.Y. Reindrawati, The Hybrid System for Recruitment and Selection During Covid-19 Pandemic, Global Business and Management Research: An International Journal, 14, 3s, 1595-1604, 2022.
14. Owolabi O.R., F.O. Olatoye, O.A. Elufioye, B. Okunade, Human resources management in healthcare: recruitment, retention, and workforce development: A review, World Journal of Advanced Research and Reviews, 21, 02, 950-957, 2024.
15. Rokhani S., Development of an Effective and Efficient Employee Recruitment and Selection System, Indonesian Journal of Innovation Multidisipliner Research, 2, 4, 247-259, 2024. Oman Z.U., A. Siddiqua, R. Noorain, Artificial Intelligence and its ability to reduce recruitment bias, World Journal of Advanced Research and Reviews, 24, 01, 551-564, 2024.

16. Oman Z.U., A. Siddiqua, R. Noorain, Artificial Intelligence and its ability to reduce recruitment bias, World Journal of Advanced Research and Reviews, 24, 01, 551-564, 2024.
17. Balcioğlu Y.S., M. Artar, Artificial intelligence in employee recruitment, Global Business and Organizational Excellence, 43, 5, 56-66, 2024.
18. Siddik A.S.A., T. Soewignyo, D.W. Mandagi, Click, Connect, Recruit: A Systematic Review of the Role of Social Media in Employee Recruitment, INFEB Jurnal Informatika Ekonomi Bisnis, 6, 2, 344-349, 2024.
19. Kostrzewska M., RECRUITMENT 5.0: THE IMPACT OF THE SOCIAL MEDIA CHANNEL LINKEDIN ON COMMUNICATION BETWEEN RECRUITERS AND CANDIDATES, Scientific Papers of Silesian University of Technology, Organization and Management Series, 199, 289-300, 2024.
20. Sýkorová Z., D. Hague, O. Dvouletý, D.A. Procházka, Incorporating artificial intelligence (AI) into recruitment processes: ethical considerations, Vilakshan-XIMB Journal of Management, 21, 2, 293-307, 2024.
21. Atluri A., V. Reddy, Cognitive HR Management: How Oracle HCM is Reinventing Talent Acquisition through AI, International Journal of Artificial Intelligence, Data Science, and Machine Learning, 6, 1, 85-94, 2025.
22. Qin X., Y.N. Huang, Z. Hu, K. Chen, L. Li, R.S. Wang, B.L. Wang, Human resource management research in healthcare: a big data bibliometric study, Human Resources for Health, 21, 94, 1-11, 2023.
23. Matthes N., T. Willem, A. Buyx, B.M. Zimmermann, Social Media Recruitment as a Potential Trigger for Vulnerability: Multistakeholder Interview Study, JMIR Human Factors, 11, e52448, 1-11, 2024.
24. Fisher E., R.S. Thomas, M.K. Higgins, C.J. Williams, I. Choi, L.A. McCauley, Finding the right candidate: Developing hiring guidelines for screening applicants for clinical research coordinator positions, Journal of Clinical and Translational Science, 6, e20, 1-7, 2021.
25. Muduli A., J.J. Trivedi, Social media recruitment: the role of credibility and satisfaction, Evidence-based HRM: a Global Forum for Empirical Scholarship, 8, 2, 237-251, 2020.
26. Hewage A., Exploring the Applicability of Artificial Intelligence in Recruitment and Selection Processes: A Focus on the Recruitment Phase, Journal of Human Resource and Sustainability Studies, 11, 3, 603-634, 2023.
27. Potočnik K., N.R. Anderson, M. Born, M. Kleinmann, I. Nikolaou, Paving the way for research in recruitment and selection: Recent developments, challenges and future opportunities, European Journal of Work and Organizational Psychology, 30, 2, 159-174, 2021.
28. Kostrzewska M., RECRUITMENT 5.0: THE IMPACT OF THE SOCIAL MEDIA CHANNEL LINKEDIN ON COMMUNICATION BETWEEN RECRUITERS AND CANDIDATES, Scientific Papers of Silesian University of Technology, Organization and Management Series, 199, 289-300, 2024.
29. Sýkorová Z., D. Hague, O. Dvouletý, D.A. Procházka, Incorporating artificial intelligence (AI) into recruitment processes: ethical considerations, Vilakshan-XIMB Journal of Management, 21, 2, 293-307, 2024.
30. Siddik A.S.A., T. Soewignyo, D.W. Mandagi, Click, Connect, Recruit: A Systematic Review of the Role of Social Media in Employee Recruitment, INFEB Jurnal Informatika Ekonomi Bisnis, 6, 2, 344-349, 2024.
31. Balcioğlu Y.S., M. Artar, Artificial intelligence in employee recruitment, Global Business and Organizational Excellence, 43, 5, 56-66, 2024.
32. Amaliyah, F.R.D. Cahyo, D.Y. Reindrawati, The Hybrid System for Recruitment and Selection During Covid-19 Pandemic, Global Business and Management Research: An International Journal, 14, 3s, 1595-1604, 2022.
33. Rokhani S., Development of an Effective and Efficient Employee Recruitment and Selection System, Indonesian Journal of Innovation Multidisipliner Research, 2, 4, 247-259, 2024.
34. Edo-Osagie E.O., J.O. Ejechi, BRIDGING DIVERSITY: INCLUSIVE HIRING STRATEGIES AND ORGANISATIONAL STRUCTURE, UNIBEN, Journal of Human Resource Management, 3, 1, 60-74, 2024.
35. Owolabi O.R., F.O. Olatoye, O.A. Elufioye, B. Okunade, Human resources management in healthcare: recruitment, retention, and workforce development: A review, World Journal of Advanced Research and Reviews, 21, 02, 950-957, 2024.

Information Security Audit – an Objective Prerequisite for Improving the Cybersecurity of the Enterprise

Rositsa Ivanova

University of National and World Economy, Sofia, Bulgaria
RIvanova@unwe.bg

Abstract. The issue of auditing the information security of the enterprise is relevant at all phases and stages of its development. Improving information connectivity between business entities operating both in the country's economy and internationally, the use of various complex digital calculations in the operational and strategic management of the business, storing and accessing data in the cloud, and others, are objective prerequisites for cyber-attacks, various frauds, data theft, as well as for the emergence of various other short-term and long-term risks. The object of consideration is the information security audit system of the enterprise, and the subject - the elements of this system. The goal is to improve this system based on an analysis of the elements of the information security audit system, giving priority to cyber risk, its inclusion in the risk management system and assessment of the enterprise's cyber security.

Keywords. Audit, information security, analysis, cyber risk, cyber security

1. Introduction

The information security audit provides objective information on the effectiveness of the measures and controls developed and implemented by the enterprise against existing, emerging and potential threats. It should prove and confirm, on the one hand, that the main risks to the enterprise are identified, monitored, analyzed, controlled and managed, and on the other hand – that managers and employees responsible for IT security have the ability to recognize and respond promptly and adequately to new threats and risks.

The information from the analysis of the discrepancies found during the audit with the generally accepted guidelines and good practices in the field of information security of the enterprise is the basis for the development and taking of correct and timely decisions for both preventive and subsequent corrective actions. Based on the information about the gaps and vulnerabilities discovered during the information security audit, the enterprise has the opportunity to obtain better protection of its main assets, as in modern conditions both the significance and the value and the relative share of long-term intangible assets are increasing.

The purpose of an independent information security audit is to demonstrate to what extent the operational and strategic decisions, actions and specific methods and means of protecting the enterprise's data meet the objectives set by its leadership and management. A well-conducted information security audit helps to add value and increase the trust of parties external to the enterprise, while ensuring compliance with regulatory requirements and information security standards.

An information security audit involves an in-depth analysis of the physical and logical protection of assets and data, administration and access levels, information security policies, as well as ensuring business continuity and planning for actions in emergencies, disasters and accidents. At the time of the information security audit, the current state of the enterprise regarding data protection is assessed.

The audit methodology is based on a risk-based approach in accordance with the requirements of the International Professional Association for Information Systems Audit and Control (ISACA), whose activities are focused on the audit and assessment of information security of enterprises, using good practices for information security. The audit of the information security of the enterprise must meet the requirements specified in the standards for this: ISO/IEC 27001 (standard for information security management systems, which defines a model for building and requirements for the specified systems); PCI DSS (a set of standards for data security in enterprises carrying out payment card activities - accepting, processing, storing or transmitting payment card information, ensuring the protection of cardholder data from fraud and violations); COBIT (an IT governance and security framework that helps enterprises manage and control their information technology, aligning it with their business objectives. The framework is a system of best practices, principles, and tools to improve risk management, compliance with regulatory requirements, and assurance of the quality and reliability of enterprise information systems); NIST (National Institute of Standards and Technology - an agency of the US Department of Commerce whose mission is to promote American innovation and the industrial competitiveness of enterprises); OWASP (an online community that publishes open source information and resources on the Internet of Things, systemsoftware, and web application security), etc.

The information security audit, carried out in accordance with the requirements of international standards for this, is a guarantee of good management and ensures high competitiveness of the enterprise in the field of its activity in which it operates and will continue to operate as a going concern.

2. Nature of the audit and types of audits

In general, an audit can be defined as an independent examination and assessment of a person, organization (enterprise), system, process, project or product in order to determine whether it meets certain criteria or standards.

According to P. Iliev, "an audit is a set of all actions and control procedures performed by authorized control bodies aimed at collecting and analysing financial and non-financial information to assess the management of financial resources and reporting in the audited entity with a view to their improvement." [8, p.2]

In the International Standards on Auditing (ISA), the nature of an audit is defined as a systematic process in which a competent independent expert collects, analyses and evaluates information that is amenable to quantitative and qualitative assessment in order to express an opinion on the degree of compliance with previously determined criteria [7].

There are different types of audits. As P. Iliev points out, “depending on the pursued goals, tasks, objects, scope and bodies of implementation, there are rich in types and diversity of audits.” [8, p.5] Due to the great diversity of audits, it is necessary to classify them according to appropriately selected criteria. For example:

A. First criterion - depending on the relationship between the subject and the object of the audit (existence of an employment-legal relationship between the enterprise and the entity performing the audit): internal and external audit.

B. Second criterion - depending on the nature and content of the audit: financial audit (an opinion is given on the reliability of the financial statements of the enterprises); performance audit; systems audit; compliance audit; audit of information systems and technologies; audit of the enterprise’s information security; social audit; marketing audit; certification audit (a consistent, documented examination of the activities, processes and systems of the enterprise in order to ensure, through clear evidence, compliance with all

requirements set by the relevant ISO standard. The purpose of the audit is to obtain a valid certificate under a given Management System); energy audit (a detailed review of the energy consumption profile of a building or group of buildings, industrial activity or installation. The energy audit is the first step towards a complete ESG transformation that turns regulatory requirements into a competitive advantage for the enterprise), etc.

C. Third criterion. Depending on the type of audited activity: public sector audit; audit of insurance companies; audit of non-profit legal entities; audit of cooperatives; audit of political parties, etc. [8, p.6]

3. Audit of the enterprise's information security

The security of the information that is processed and stored in the enterprise, as well as of its information arrays, essentially represents protection against threats and various types of risks. This creates the need for permanent analysis and assessment of the enterprise's information security. Moreover, new types of threats and risks to the security of the information that is created in the enterprise or comes from external entities, functions, is used by the enterprise and is stored in its information arrays are constantly emerging. Based on the information obtained from the analysis, the weaknesses, vulnerabilities and gaps that may exist in the information network and/or information systems of the enterprise are revealed and identified. The comprehensive audit of the security of the information and information arrays of the enterprise covers many and in-depth policies and procedures in the modern conditions of constantly emerging various threats, risks and new vulnerabilities. In these conditions, audits must verify the scope of controls to ensure stable protection of data, information, information arrays of the enterprise and in general of the enterprise's information security management system. Moreover, the sustainable development of the enterprise implies and requires the implementation of additional costs to achieve sustainability in its three components (economic, social and environmental). The costs of cybersecurity of the enterprise, reflected at the first level as costs of economic elements, and at the second level - as administrative costs, have a direct impact on the financial results, and hence on the financial condition and financial stability of the enterprise.

It is necessary to note that the assessment of vulnerabilities is not limited to technology, but it is extended to the processes and people who are part of the information security ecosystem, and can be generalized to the cybersecurity of the enterprise.

4. Internal audit

In accordance with the provisions of the ISO 27001 standard, an enterprise can undergo internal and external audits to demonstrate that it adheres to current best practices for information security management.

An ISO 27001 audit is a systematic and independent examination of an enterprise's information security management system (ISMS) to determine whether it complies with this standard. The purpose of an information security audit is to verify whether the enterprise's ISMS meets the requirements of the ISO/IEC 27001:2022 standard, the latest version of which is 2022, and on this basis to assess the effectiveness of the system and identify opportunities for its improvement.

An internal audit is a review of the enterprise's ISMS carried out by internal personnel, with an employment relationship between the audited enterprise and the personnel performing the audit. It is possible for an internal audit to be performed by an external contractor hired to work with an internal team, in which case the audit of the enterprise's information security is considered an internal audit. Even when an internal

audit is performed by an external party, it is considered an internal audit. The exception to this is when the external auditor is part of an ISO 27001 certification body, and the audit is classified as an external information security audit.

The internal audit of the enterprise's information security management system (ISMS) covers the following stages:

1) a review of the current IT processes and the scope of the enterprise's ISMS;

2) a review of the internal documentation regarding the ISMS policies and procedures, as well as the organization for its maintenance in the enterprise;

3) an analysis and assessment of the risk and weaknesses of the enterprise's ISMS;

4) review and study of evidence from the enterprise's activities that identify whether the policies and procedures of the ISMS are applied in their logical sequence and interconnectedness;

5) comparative analysis of the information from the review of internal documents on the ISMS policies and procedures and the requirements contained in the ISO 27001:2022 standards;

6) based on the findings made during the information security audit, development and implementation of improvements regarding its ISMS in the enterprise's activities.

The enterprise should develop and consistently implement an internal audit program in order to maintain compliance with the requirements of the ISO 27001 standard (see clause 9.2 of the standard), which should determine and record: the frequency of internal audits; the methods that will be used when conducting internal audits; the persons responsible for planning, conducting, completing and reporting the results of internal audits of the enterprise's ISMS. We consider it appropriate to point out that it is appropriate to conduct an internal audit of the ISMS at least once a year, for example, towards the end of the current reporting period, which should be tied to the preparation of the company's annual financial statements.

5. External Audit

On the issue of external audit of the enterprise's information security, we will focus our attention on three important areas:

- certification audits according to ISO 27001 standards;
- audits according to the Personal Data Protection Regulation;
- audits according to the Network and Information Security Directive.



Fig. 1. ISO Standards 27001:2022

The new ISO 27001:2022 framework (effective October 2022 – see Figure 1) adds 11 new controls that compliant enterprises must implement in their practices. Some of the new areas included in the audited controls are:

- Threat intelligence and continuous risk management;
- Cloud information security;
- Business continuity from a communications systems perspective;
- Physical security monitoring;
- Configuration management;
- Deletion of sensitive data;
- Appropriate data masking (anonymization);
- Web and content filtering;
- Use of secure encryption principles.

Enterprises interested in ISO 27001 certification should undergo four external audits:

1) ISMS design review (the auditor reviews the enterprises documentation, processes and procedures in order to express an opinion and ensure that the controls and design of the information security management system comply with ISO 27001 standards;

2) certification audit (the auditor reviews the business processes and controls in the enterprises in order to ensure that they meet the requirements of ISO 27001);

3) surveillance audits (the auditor performs periodic (surveillance) audits of specific areas of the enterprises information security management system in order to express an opinion and ensure that it implements the procedures and processes defined and recorded in its internal documents concerning this system);

4) recertification audits (carried out every three years and covering all areas of the enterprise's information security management system, providing guarantees that it continuously follows the requirements of the ISO 27001 standards and improves its security system when new risks arise regarding information and its information arrays.

In order to process the personal data of their employees, as well as of external entities related to their activities, enterprises must have a legal basis for this. They should develop and implement in their activities a system of actions and processes that will become a quality and effective audit model under the General Data Protection Regulation. At the same time, each enterprise should organize continuous monitoring and improvement of the audit model by developing a system of appropriately selected measurable indicators that lend themselves to maturity modelling. The above gives rise to the significance, importance and relevance of the issue of auditing under the General Data Protection Regulation (see Figure 2).



Fig. 2. GDPR Directive - General Data Protection Regulation

The audit under the data protection regulation provides assurance to both the enterprise itself and its business partners, as well as to the state and municipal administration, that it has complied and continues to comply with the requirements of the regulation on the protection of personal data of individuals with regard to their processing and their free movement.

The regulation also applies to entities located outside the European Union and the European Economic Area, but which process data of citizens of the European Union member states. The main objective of the general regulation is to guarantee to citizens that, in implementing continuous and effective control over their personal data, both compliance gaps and risks (current and potential) for the enterprise are permanently identified, by developing and implementing appropriate measures to eliminate them in its activities.

On 16 January 2023, the Network and Information Security 2 (NIS 2) Directive on measures to achieve a high common level of cybersecurity in the European Union (NIS2 Directive) and the Directive on the Resilience of Critical Entities (CER Directive) entered into force. The NIS2 Directive repeals the existing NIS Directive, creating a more comprehensive and harmonised system of cybersecurity rules for enterprises operating in the Member States of the European Union. The CER Directive repeals the European Critical Infrastructure Directive and introduces new, stricter rules for the cyber and physical resilience of critical companies, structures and networks. The NIS2 Directive is the first piece of pan-European legislation on cybersecurity for enterprises operating within the European Union. The aim of the NIS2 Directive is to achieve a high common level of cybersecurity in the Member States of the European Union.

H. Stanev points out that "a study for the period June 2021 - June 2022, cited by the European Union Agency for Cybersecurity (ENISA), identifies eight main threats to cybersecurity in the EU:

- Ransomware - taking control of data on a given device with a subsequent request to pay a certain amount in order to restore access to the data;

- Malware - through computer viruses, worms, Trojan horses and spyware, a system is penetrated to be used for other purposes (e.g. cryptojacking (secret use of processing power to create cryptocurrencies) or gaining control over cameras or routers);

- Exploiting human errors - about 60% of system breaches in Europe, the Middle East and Africa (Internet address from https://www.europarl.europa.eu/topics/bg/article/20220120STO21428/kibersighurnost-ghlavnite-i-novite-zaplakhi) are related to the use of human errors by misleading users and opening a dangerous file or visiting an insecure website, unauthorized access to computer systems or data is provided. Most often, such attacks begin with sending an email ("phishing") or text messages ("smishing") that appear to be sent by official financial institutions (banks);

- Threats to data (Threats to data) - methods such as Ransomware, Malware, Phishing, breaches (intentional attacks) or leaks (unintentional release) of data are used;

- Threats to the availability of data or services interrupting users' access to their information resources by overloading the network infrastructure or damaging the system (so-called Denial of Service (DoS) attacks, Distributed Denial of Service (DDoS) attacks, Ransomware or Hardware failures);

- Threats to internet access (Threats to internet access) - taking control or destroying the internet connectivity infrastructure (again through Denial of Service (DoS) attacks, Distributed Denial of Service (DDoS) attacks, Network outages or Government censorship);

- Disinformation and propaganda (Disinformation and propaganda) by creating fake images, videos or audio recordings that are difficult to distinguish from the real ones (e.g. bots impersonate real public figures and send numerous fake comments);

- Supply chain attacks - interference in business relationships between suppliers and customers, by simultaneously attacking to obtain economic benefit (inviting payment on a real invoice but on the account of a third party) or attempting to worsen the relationship between them (so-called Inserting malicious code, compromising third-party services or Exploiting vulnerabilities)." [9, p.1041]

The cybersecurity audit should focus on the basic rules that the enterprise should apply during its activities in order to build sustainable protection against modern threats, which are showing a tendency to diversify, deepen and become more difficult to recognize. These rules can be reduced to:

1) strengthening systems by eliminating unnecessary services, which aims to reduce the possibilities for cyberattacks;

2) access management - applying the principle of least privilege and two-factor authentication;

3) network protection - segmentation, TLS inspection, providing important improvements such as improved speed and encryption, as well as data traffic monitoring;

4) constant real-time monitoring and incident analysis using AI-based tools;

5) sustainable archiving - regularly testing and maintaining backups located outside the main networks.

The audit stages in accordance with the Network and Information Security Directive 2 (NIS 2) can be presented in the following sequence and relationship between them:

1) assessment of the degree of usefulness of network and information security measures;

2) identification of weak points and assessment of the riskiest ones;

3) implementation of improvements, with information about this being disclosed in a timely manner;

4) linking the indicators for measuring cybersecurity with real business indicators, such as the degree of vulnerability, response time, security improvement;

5) verification of privileged accounts and protection of administrator and accessible profiles;

6) verification of the personnel who introduce and promote the use of good practices in each department of the enterprise;

7) verification of the current state of the enterprise, the place it occupies in terms of network and information security and establishment of the necessary actions to achieve full sustainability of the protection of the enterprise's information and information arrays;

8) organizing constant monitoring and vulnerability scanning;

9) building a system for regular vulnerability scanning, access management and backup storage in order to achieve sustainable protection.

In our opinion, an enterprise's cybersecurity issues should be integrated into its mission, vision, and strategy (see Figure 3).



Fig. 3. Integrating information security auditing into enterprise strategy

6. Stages of the enterprise information security audit

The enterprise information security audit is a system of elements between which there are objective relationships and dependencies. One of the elements of this system is the stages in conducting the information security audit and the enterprise information security management system. The stages of the audit can be systematized in the following sequence: preparatory stage; implementation; reporting; results.

1) First stage. Preparatory stage.

During the preparatory stage, the following activities are carried out:

Activity 1. Audit planning.

When planning the audit, its goals and objectives are determined, through the solution of which the goals can be achieved. The goals and objectives of the audit should be defined in an appropriate manner, be clearly and precisely formulated, be understandable, measurable, achievable and lead to the achievement of the results desired by the management. They should cover all the main areas of the enterprise's information and technical infrastructure. For example, the objectives of the audit may include compliance with regulatory requirements, including those for cybersecurity, assessment of network security, disclosure of potential weaknesses and vulnerabilities in the software and hardware used by the enterprise, and others. Also, when planning the audit, training of the enterprise's employees (staff) should be carried out on best practices in the field of cybersecurity and the protection of its information arrays and systems. In conclusion, it should be noted that the prepared plan for conducting an IT security audit should be approved by the company's management.

Activity 2. Formation of an audit team.

This activity involves the formation of an audit team of highly qualified and experienced professionals in various aspects of IT security, such as network security specialists, system administrators and compliance experts, etc. During this activity, a preliminary meeting of the audit team should be organized, at which it should be introduced and the main strategies, goals and objectives adopted by the enterprise in the field of IT security should be discussed. At this meeting, the tasks, responsibilities and deadlines for implementation should be distributed to each of the members of the audit team.

Activity 3. Audit Information Assurance.

This activity is related to the collection of the necessary documentation, formation and description of the enterprise's policies, concerning and related to the enterprise's IT infrastructure. These include network diagrams, system configurations, security policies,

access control and incident response plans. The information contained in the enterprise's current documentation and policies is of great importance in providing complete and comprehensive data on the state of its IT security. Based on this information (audit information assurance), the audit team has the opportunity to conduct an in-depth analysis and assessment of the security of the enterprise's IT infrastructure.

Activity 4. Benchmarking and assessing compliance with security standards and regulations.

This activity should assess the compliance of the enterprise's policies with applicable standards and regulations in the field of IT security. Failure to comply with the requirements can lead to large fines, damage to the reputation of the enterprise and deterioration of its financial results, condition, stability and competitiveness. During the IT security audit, compliance with regulatory requirements in this area should be assessed, as well as the possible potential risks associated with their non-compliance.

Activity 5. Conduct Vulnerability Assessments and Penetration Testing.

Vulnerability assessments and penetration testing help auditors uncover weaknesses in networks, systems, and applications. Vulnerability assessments involve scanning IT infrastructure to detect security gaps, while penetration testing simulates real-world cyberattacks. The goal is to provide an assessment of an enterprise overall security resilience and security controls, providing valuable insight into areas that require improvement in their security posture.

Activity 6. Analysis of access control and user permissions.

This activity includes a detailed review, permanent and comprehensive control of access and user permissions for access to user information systems. The goal here is to determine whether only authorized persons have access to sensitive information in the enterprise. During this activity, it should be determined whether the principle of least privilege is observed, in accordance with which the enterprise should provide users with the minimum level of access that is necessary and sufficient to perform their job functions. Along with this, it is necessary to assess the enterprise's policies for maintaining passwords and their timely updating, as well as multi-factor authentication and processes for revoking access to information systems, with which the enterprise is able to protect itself from unauthorized access. Therefore, activity's 6 are aimed at reviewing, critically analysing and optimizing access control measures to information arrays and systems in order to ensure reliable security for them.

Activity 7. Review of physical security measures.

Observations in practice show that in most enterprises the issue of physical security of information arrays and systems is neglected, even when conducting an IT security audit. It should be noted that the review of physical security measures is important and they are crucial for protecting sensitive equipment and information from unauthorized access, theft and damage. The activities here include a thorough and comprehensive review of the security measures implemented on the premises of the enterprise, including access control systems, security cameras and protected server rooms.

Activity 8. Assessment of incident response and disaster recovery plans.

During this activity, a review of the enterprise's incident response and disaster recovery plans is carried out as part of the IT security audit. A well-prepared response plan allows to minimize potential impacts and breaches in the enterprise's IT security. At the same time, a reliable disaster recovery plan helps and enables the rapid restoration of the enterprise's operations after data loss. Particularly important in modern conditions is the permanent and

reliable assessment of the enterprise's existing plans in order to promptly identify areas for improvement, to recommend the use of best practices, by adopting quick and effective measures for adequate responses to incidents related to the enterprise's cybersecurity.

Activity 9. Identification, analysis and assessment of the risk associated with the main activities and assets (resources) of the enterprise, on the basis of which the IT security audit will be conducted.

2) ***Stage 2. Execution.***

During the implementation of the enterprise's IT security audit, the following activities are performed:

Activity 1. Assessment of the effectiveness of the controls used in the enterprise to achieve information security.

Activity 2. Assessment of the enterprise's communication and network security.

Activity 3. Assessment of the security of the main and web-based applications used by the enterprise during its activities.

Activity 4. Assessment of the security of the enterprise's databases.

Activity 5. Assessment of the security of the main server complexes and the services provided by them in carrying out the enterprise's business.

Activity 6. Analysis and assessment of the methods used for security of endpoints and individual workstations in the enterprise.

Activity 7. Scanning for vulnerabilities and potential threats to the security of information of the Internet-accessible systems and applications that the enterprise uses in carrying out its business.

3) ***Stage 3. Reporting.*** During this stage, a comprehensive and final report is prepared and presented, covering all findings made during the audit of the enterprise's IT security, including the identified possible measures to improve the security of its information arrays and systems.

4) ***Stage 4. Result.*** The final report on the audit of the enterprise's IT security includes a detailed description of the conclusions after the analyses and assessments of security on the individual systems. As a supplement to the report, recommendations for improving security and optimizing the enterprise's information environment are also indicated.

7. Conclusion

The benefits for the enterprise from a fully performed IT audit of its information security can be systematized in the following directions:

1. A competent and qualitatively prepared analysis of the current state of the enterprise's information security provides the basis on which to plan and ensure better and timely active protection of its key information assets.

2. The recommendations made during the audit and the application of a systematic approach allow for the identification and prioritization of the necessary changes in the design, planning and budgeting of the enterprise's information infrastructure activities.

3. Based on the final report from the audit, a specific work program is drawn up to improve information security, taking into account the gaps and vulnerabilities found in the communication and information systems used by the enterprise.

4. The audit creates the prerequisites for achieving fuller compliance with both good practices and regulatory requirements in the field of information security of the enterprise.

5. The actions taken to eliminate the gaps identified by the audit regarding the enterprise's information security help maintain more reliable communication with all parties interested in its activities, as well as improve security when working with partners and other enterprises and organizations.

References

1. Directive (EU) 2016/1148 of the European Parliament and of the Council concerning measures for a high common level of security of network and information systems.
2. Directive (EU) 2022/2555 (NIS 2 Directive) of the European Parliament and of the Council of 14 December 2022 concerning measures for a high common level of cybersecurity in the Union, amending Regulation (EU) No 910/2014 and Directive (EU) 2018/1972 and repealing Directive (EU) 2016/1148
3. Concerning measures for a high common level of cybersecurity in the Union
4. Regulation on minimum requirements for network and information security, adopted by Council of Ministers No 186/26.07.2019, published in the State Gazette, issue 59/26.07.2019
5. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)
6. Bankova, D. (2020). On cybersecurity in enterprises, electronic magazine of IDES, issue 1, pp. 1-14.
7. Veysel, A. (2015). The audit process according to the International Standards on Auditing, electronic magazine of IDES, issue 1, pp. 1-15.
8. Iliev, P. (2019). Current issues in the implementation of auditing, NBU Internet address: https://eprints.nbu.bg/id/eprint/4123/1/%D0%A1%D1%82%D1%83%D0%B4%D0%B8%D1%8F%20%D0%BE%D0%B4%D0%B8%D1%82%D0%B8%D0%BD%D0%B3.pdf)
9. Staneva, V., H. Stanev (2024). On some aspects of the application of the Cybersecurity Directive in the accounting of micro and small enterprises in Bulgaria. Forty-seventh International Scientific Conference Knowledge in practice, 13-15.12.2024, Bansko, IKM, Skopje, Vol. 67.6., p. 1039-1045.
10. Stanev, H. (2020). Prerequisites for sustainable development of aviation in the Republic of Bulgaria. Anniversary scientific conference on the occasion of 70 years since the establishment of the department and specialty "Economics of Transport" on the topic "Transport Connectivity 2020", UNWE, Sofia, 29.10.2020, pp. 94 – 100.
11. Internet source: https://e-security.bg/5-klyuchovi-navika-za-kiberhigiena-koito-mogat-da-spasyat-milioni/

Economic Metrics for Measuring the Effectiveness and Efficiency of Cybersecurity Investments

Nedko Tagarev

University of National and World Economy, Sofia, Bulgaria
ntagarev@unwe.bg

Abstract. In an era of escalating cyber threats, organizations face the persistent challenge of justifying and optimizing financial allocations toward cybersecurity. This paper addresses this challenge by presenting a comprehensive framework of economic metrics designed to measure the effectiveness and efficiency of cybersecurity investments quantitatively. Moving beyond technical assessments, the study grounds itself in financial and risk-based models to provide a business-centric evaluation of security initiatives. The methodology systematically outlines key quantitative tools, including Total Cost of Ownership (TCO) for assessing long-term investment value, Single Loss Expectancy (SLE) and Annualized Loss Expectancy (ALE) for quantifying cyber risk in monetary terms, and Return on Security Investment (ROSI) for evaluating the financial return of security controls. Furthermore, the analysis incorporates capital budgeting techniques such as Net Present Value (NPV) to account for the time value of money in multi-year security projects. The results demonstrate that the integrated application of these metrics enables organizations to transition from qualitative security assessments to data-driven decision-making. This allows for the direct comparison of alternative security solutions, the objective justification of security budgets to stakeholders, and the strategic alignment of cybersecurity posture with overarching business objectives. The discussion acknowledges the inherent difficulties in estimating parameters, such as threat probability, but concludes that a disciplined, metric-driven approach is indispensable for rational resource allocation. By adopting this economic perspective, organizations can not only enhance their security resilience but also demonstrate tangible value and fiscal responsibility, thereby transforming cybersecurity from a perceived cost center into a strategically managed business function.

Keywords: Cybersecurity Economics, Return on Security Investment (ROSI), Annualized Loss Expectancy (ALE), Total Cost of Ownership (TCO), Net Present Value (NPV), Risk Quantification, Security Metrics, Cybersecurity Investment, Economic Efficiency, Cyber Risk Management.

1. Introduction

!!!Disclaimer: This publication does not constitute legal or economic advice. The values used are hypothetical and may vary depending on the object, subject matter, and context, as well as the possible values of different service providers. It should be noted that the result will not always be positive, as in the scenario shown.!!!

In the contemporary digital landscape, characterized by pervasive connectivity and an ever-expanding attack surface, cybersecurity has transcended its traditional role as a technical IT function to become a cornerstone of organizational resilience, economic stability, and strategic governance. The relentless evolution of cyber threats - from financially motivated ransomware syndicates to state-sponsored espionage campaigns - poses a clear and present danger to businesses, governments, and critical infrastructure worldwide. As organizations

increasingly digitize their operations and assets, the potential financial, operational, and reputational damage from cyber incidents has reached unprecedented levels. The global economic impact of cybercrime is projected to soar into the trillions of dollars, underscoring a systemic risk that demands a proactive and strategically aligned response.

However, despite this universal recognition of cyber risk, Chief Information Security Officers (CISOs) and security leaders continue to grapple with a fundamental and persistent challenge: justifying cybersecurity budgets and demonstrating their value to executive leadership and board members. Cybersecurity investments are often perceived as a "cost center"—a necessary expense with an intangible return, rather than a value-driven investment that contributes positively to the organization's strategic objectives. This perception gap stems from the inherent difficulty in quantifying what did not happen due to successful security interventions. Unlike a new sales platform that can be measured by revenue growth, the success of a security control is often measured by the absence of loss, creating a significant hurdle for effective communication and resource allocation.

This communication gap is frequently bridged with technical jargon and fear-based appeals, which are increasingly ineffective in a competitive business environment defined by rigorous financial scrutiny and accountability. Executive decision-makers, trained to evaluate investments based on Return on Investment (ROI) [1], Net Present Value (NPV) [2], and Total Cost of Ownership (TCO) [3], require a financial and economic lexicon to understand and approve security initiatives. The inability to translate technical security postures into business-relevant financial terms can lead to underinvestment, the misallocation of resources, and an overall security posture that is misaligned with the organization's actual risk appetite and business strategy.

To address this critical gap, the field of cybersecurity economics has emerged as a vital discipline. It provides a framework for applying well-established economic principles and financial metrics to the domain of information security. This paradigm shift moves the conversation from "how secure are we?" to "how much security is economically efficient?" and "what is the optimal level of investment in cybersecurity?". It acknowledges that perfect security is an unattainable and economically irrational goal; instead, the objective is to manage cyber risk to a level that is acceptable to the business, ensuring that every dollar spent on security delivers the maximum possible reduction in expected loss.

This paper aims to contribute to the evolving discipline by synthesizing and presenting a comprehensive framework of key economic metrics specifically designed to evaluate the effectiveness and efficiency of cybersecurity investments. The primary objective is to equip security professionals and decision-makers with a practical, quantitative toolkit for strategic planning and justification. The research question central to this work is: How can a defined set of economic metrics be systematically applied to measure and communicate the value and performance of cybersecurity programs?

To answer this question, this paper is structured as follows. The subsequent Methods section will provide a detailed exposition of the core economic models, including Total Cost (TC) for program budgeting, Total Cost of Ownership (TCO) for lifecycle analysis, and risk quantification metrics such as Single Loss Expectancy (SLE) and Annualized Loss Expectancy (ALE). It will further delve into investment appraisal techniques, namely Return on Security Investment (ROSI) and Net Present Value (NPV), which are crucial for comparing alternative security solutions and justifying capital expenditures. The Results section will illustrate the practical application of these metrics, demonstrating how they can be used to calculate risk savings, net savings, and ultimately, a financially sound ROI for security projects. The Discussion will address the practical challenges and limitations of these models, particularly concerning data quality and estimation uncertainty, while arguing for their indispensable role

in modern, risk-based cybersecurity management. Finally, the Conclusion will summarize the findings and reaffirm the necessity of integrating economic analysis into the core of cybersecurity governance, thereby transforming cybersecurity from a technical safeguard into a strategically managed business enabler.

2. Methodology

The following economic metrics were analyzed based on established financial and risk assessment methodologies.

2.1 Total Cost (TC)

The total cost of a cybersecurity program as a project can be defined as:

$$TC = I_{new} + OC_{new} + OC_{call} \dots \quad (1)$$

where,
TC – total costs;
I_{new} - research and investment costs,
OC_{new} - ongoing operational costs;
OC_{call} - the costs of maintaining legacy systems until their retirement [4].

2.2 Annualized Loss Expectancy (ALE)

ALE estimates the expected loss from an asset over one year and is calculated as [5]:

$$ALE = ARO \times SLE \dots \quad (2)$$

where,
ALE - expected loss from an asset over one year;
ARO- Annual Rate of Occurrence;
SLE - Single Loss Expectancy.
"Annualized Rate of Occurrence, also known as ARO, refers to the expected frequency with which a risk or a threat is expected to occur. ARO is also commonly referred to as Probability Determination." [6].
"Single Loss Expectancy (SLE) is defined as the cost of a single loss, calculated by multiplying the Asset Value (AV) by the Exposure Factor (EF)." [7].

2.3 Single Loss Expectancy (SLE)

SLE quantifies the monetary loss expected per risk occurrence:

$$SLE = AV \times EF \dots \quad (3)$$

where,
AV - the Asset Value;
EF - Exposure Factor.

2.4 Total Cost of Ownership (TCO)

TCO includes both the purchase price and operational costs of an asset over its lifecycle. It provides a long-term perspective on the value of security investments [8].

2.5 Return on Security Investment (ROSI)

ROSI measures the financial return of security investments by comparing the cost of the investment to the risk reduction achieved:

$$ROSI = \frac{\text{Risk Savings} - \text{Security Investment}}{\text{Security Investment}} \times 100\% \ldots \quad (4)$$

2.6 Net Present Value (NPV)

NPV can evaluate the profitability of a security investment over time:

$$\text{NPV} = \sum_{t=1}^{T} \frac{C_t}{(1+r)^t} - C_0 \ldots \quad (5)$$

where,

C_t -net cash flow during period;

C_0 - the initial investment;

r -the discount rate;

T - number of periods.

2.7 Expected Annual Risk Cost

This metric aggregates the expected cost of all threats

$$\text{Expected Annual Risk Cost} = \sum(\text{Threat Probability} \times \text{Potential Loss}) \ldots \quad (6)$$

2.8 Risk Savings and Net Savings

Risk Savings are calculated as:

$$\text{Risk Savings} = \text{Total Base Risk} \times \text{Risk Reduction Rate} \ldots \quad (7)$$

Net Savings are derived from:

$$\text{Net Savings} = \text{Total Risk Savings} - \text{Total Cost} \ldots \quad (8)$$

2.9 ROI Based on Net Savings

The final ROI is calculated:

$$\text{ROI} = \left(\frac{\text{Net Savings}}{\text{Total Cost}}\right) \text{x } 100\% \ldots \quad (9)$$

3. Results

This section presents a comprehensive application of the economic metrics outlined in the Methods section to a hypothetical scenario with the implementation of a Security Information and Event Management (SIEM) system. The analysis will demonstrate the step-by-step calculation of each metric and its interpretation for decision-making.

3.1 Scenario definition and Base assumptions

The scenario: *A company processes sensitive customer data and is subject to stringent regulatory requirements. The existing security infrastructure lacks centralized monitoring, leading to a high risk of undetected data breaches. The proposed SIEM solution is projected to have a lifespan of five years.* For the scenario, we have two hypotheses: 1) Assumption without SIEM and 2) Assumption with SIEM.

The first step **is to define values for Base Risk Assumptions (Without the SIEM). "Risk Assumption is a risk management tool where a company chooses to make itself responsible for potential losses, such as not obtaining insurance due to minimal risk of loss or opting for self-insurance by setting aside funds for future losses."** [9]

- **Asset Value (AV) of Protected Data:** $5,000,000;
- **Threat:** Major data breach via an undetected advanced persistent threat (APT). An advanced persistent threat (APT) is a sophisticated, sustained cyberattack in which an intruder establishes an undetected presence in a network to steal sensitive data over a prolonged period. An APT attack is carefully planned and designed to infiltrate a specific organization, evade existing security measures, and fly under the radar."[10]
- **Exposure Factor (EF):** 40% (estimation of data loss and exposure);
- **Single Loss Expectancy (SLE):**
  - **Annual Rate of Occurrence (ARO):** 20% (a 1-in-5-year probability of a significant breach);
  - **Annualized Loss Expectancy (ALE - Base):** This $400,000 represents the expected annual cost from this specific threat in the current, unimproved state.

***The second step* is to define values for Investment Cost Assumptions (For the New SIEM)." Security information and event management, or SIEM, is a security solution that helps organizations recognize and address potential security threats and vulnerabilities before they have a chance to disrupt business operations."**[11]

- **Initial Setup Cost (I_{new}):** $150,000 (includes software licenses, hardware, and initial configuration).
- **Annual Operational Cost (OC_{new}):** $50,000 (includes maintenance, tuning, and dedicated analyst time).
- **Project Lifespan (n):** 5 years.
- **Discount Rate (r):** 8% (reflecting the company's cost of capital)." Cost of capital is the minimum rate of return or profit a company must earn before generating value. It's calculated by a business's accounting department to determine financial risk and whether an investment is justified."[12]

3.2 Application of Economic Metrics

Total Cost of Ownership (TCO). The TCO provides a realistic picture of the total financial commitment over the system's lifespan, moving beyond the mere purchase price. The TCO of the SIEM solution is $400,000. This figure is essential for budgeting and is a key input for subsequent ROI and NPV calculations.

Risk Reduction and Risk Savings. The primary value of the SIEM is its ability to reduce risk. It is estimated that the SIEM will enhance threat detection and response, thereby reducing the likelihood of a successful major breach. **Post-SIEM ARO is** reduced from 20% to 5%.

Post-SIEM ALE - Risk Reduction Rate is 75%. The monetary value of this risk reduction is the Risk Savings. The SIEM investment is expected to generate $300,000 in annual risk savings by significantly lowering the expected annual loss from data breaches.

Net Savings and Return on Investment (ROI). Net Savings and ROI translate risk reduction into direct financial performance indicators. Over five years, the SIEM project yields net savings of $1.1 million and an impressive ROI of 275%. This provides a powerful, easily understandable argument for the investment, demonstrating that for every dollar spent, the company expects a return of $2.75 in saved losses.

Net Present Value (NPV). While ROI is beneficial, NPV offers a more sophisticated analysis by taking into account the time value of money. It discounts future cash flows (in this case, the annual risk savings) to their present value. The annual net cash flow (C_t) is the risk savings minus the yearly operational cost: $300,000 - $50,000 = $250,000. Note that the initial cost (C_0) is $150,000. The NPV is strongly positive at $848,177. A positive NPV indicates that the projected earnings (in terms of risk savings) generated by the investment exceed the

anticipated costs, when considered in present-value terms. This makes the SIEM project a financially attractive and value-adding initiative.

Return on Security Investment (ROSI). ROSI offers a security-specific variant of ROI, focusing purely on the efficiency of the security spend itself. The annualized security investment is the TCO divided by the project lifespan: \$400,000 / 5 = \$80,000. The ROSI of 275% confirms the high efficiency of the investment from a security-centric viewpoint, aligning with the classic ROI calculation and providing a consistent narrative of substantial value.

The results are summarized in Table 1.

Table 1. Summarised results.

Metric	*Formula*	*Result*	*Interpretation*
Base ALE	SLE×ARO	\$400,000	Expected annual loss without the new SIEM.
TCO	$I_{new}+(OC_{new}\times n)$	\$400,000	Total lifecycle cost of the SIEM investment
Risk Savings	$ALE_{base}-ALE_{new}$	\$300,000/yr	Annual monetary benefit from reduced risk.
Net Savings	$(RiskSavings\times n)-TCO$	\$1,100,000	Total net financial benefit over 5 years.
ROI	$NetSavings/TCO)\times 100\%$	275%	High overall return on the invested capital.
NPV	$\sum_{t=1}^{T}\frac{C_t}{(1+r)^t}-C_0$	\$848,177	The investment adds significant present value.
ROSI	$\left(\frac{\text{Net Savings}}{\text{Total Cost}}\right)$ x 100%	275%	The security spending itself is highly efficient.

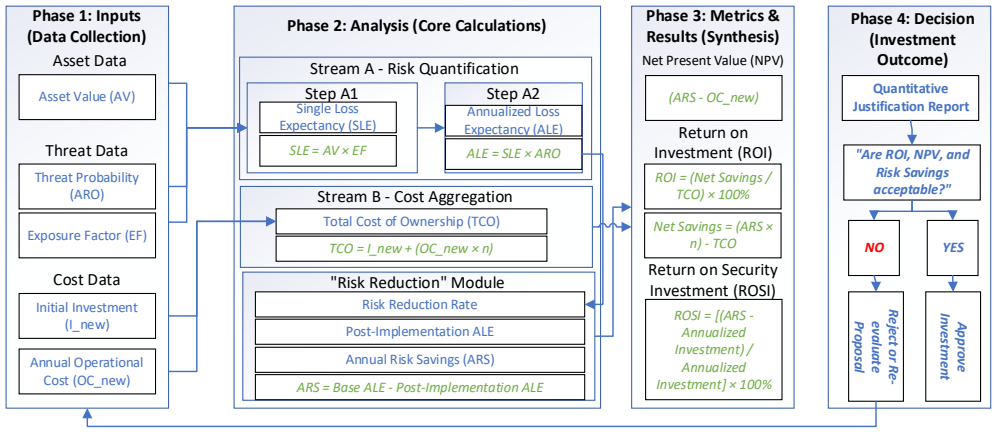
The results uniformly and strongly indicate that the proposed SIEM investment is economically justified. It transitions the justification from a qualitative, fear-based argument ("we might get hacked") to a quantitative, business-centric case, demonstrating substantial risk reduction, a positive net present value, and a high return on investment. This multi-faceted economic analysis provides management with the robust financial data needed to approve the project confidently.

4. Discussion

The comprehensive results from the SecureFin case study demonstrate the powerful utility of economic metrics in cybersecurity decision-making. The uniformly positive outcomes—a 275% ROI, \$848k NPV, and \$300k in annual risk savings—present a compelling financial case for the SIEM investment. However, **the actual value of this analysis lies not merely in the positive numbers but in the more profound insights, limitations, and strategic implications that these metrics reveal**.

This discussion interprets these results, addresses the challenges of quantitative risk assessment, and explores the organizational shift required to leverage such frameworks effectively.

Figure 1: Graphical Process Model for Cybersecurity Economic Evaluation



Interpretation of Economic Outcomes

The first and most direct interpretation of the results is that the SIEM investment is not a cost, but a value-generating activity. The substantial positive NPV indicates that the project is expected to increase the value by over $848,000. This aligns the cybersecurity initiative with capital projects in other parts of the business, such as new manufacturing equipment or market expansion, which are also evaluated on their value-creation potential. The high ROI and ROSI further reinforce this, showing that the capital allocated to security can compete with, and even outperform, other potential investments within the organization. This fundamentally reframes the conversation for the board and CFO, moving cybersecurity from an operational expense to be minimized to a strategic investment to be optimized.

Furthermore, the decomposition of value through metrics like Risk Savings ($300,000/year) and ALE provides critical granularity. It allows security leaders to move beyond vague statements about "improving security" and instead articulate a precise narrative: "This investment specifically addresses a $400,000 per year data breach risk, and we are reducing it by 75%." This level of specificity is crucial for prioritization. Suppose another proposed control, such as an endpoint detection and response (EDR) system, claims to address a different threat with an ALE of only $50,000. In that case, the organization can now make an informed, data-driven decision on where to allocate limited resources first.

The Critical Role of Assumptions and Data Quality

The robustness of this entire economic model is entirely contingent on the accuracy of its underlying assumptions. The results, while compelling, are a direct function of the estimated ARO (20% reduced to 5%), the Exposure Factor (40%), and the Asset Value ($5 million). These figures are often the most challenging to derive with high confidence.

The ARO, in particular, is notoriously difficult to pin down. It is often extrapolated from industry reports, historical internal incident data (which may be incomplete), or expert opinion. A slight change in this assumption can dramatically alter the outcome. For instance, if the base ARO were only 10% instead of 20%, the base ALE would be $200,000, and the annual risk savings would drop to $150,000. This would lower the NPV to approximately $299,000 and the ROI to 87.5%, still positive but significantly less compelling. This sensitivity underscores the necessity of conducting a sensitivity analysis or Monte Carlo simulation alongside the primary calculation to present a range of possible outcomes, rather than a single, potentially fragile, number.

Similarly, the Asset Value (AV) must reflect more than just the replacement cost of hardware or software. For a data asset, it should encompass regulatory fines, reputational damage, customer churn, operational downtime, and legal fees. Quantifying these intangible factors requires collaboration with leaders from legal, communications, and business units. The EF is another estimation that demands careful consideration; is 40% of the data asset truly exposed in a single breach? Refining these inputs through cross-functional workshops and leveraging industry frameworks, such as FAIR (Factor Analysis of Information Risk), can significantly enhance the model's credibility.

Strategic and Organizational Implications

Adopting an economic metrics framework necessitates a significant cultural shift within an organization. It requires the cybersecurity team to develop financial literacy and the finance team to develop a basic understanding of cyber risk. This collaboration is essential for validating assumptions and ensuring the model aligns with the organization's overall risk appetite and financial planning processes. The CISO's role evolves from a technical expert to a business leader who can articulate risk in the language of the boardroom.

Moreover, this approach champions the principle of **economic efficiency in cybersecurity**. It implicitly acknowledges that not all risks are worth mitigating and that the goal is optimal risk management, not risk elimination. The analysis might reveal that for specific low-ALE threats, the TCO of a proposed control exceeds the potential risk savings, making it an economically irrational investment. In such cases, the organization may rightfully choose to accept, transfer, or mitigate the risk through a less costly means. This disciplined, cost-effective approach prevents overspending on security and ensures that resources are allocated to areas of the highest business impact.

Limitations and Future Directions

While powerful, this methodology is not a panacea. Its primary limitation is its reliance on estimation and prediction. It cannot account for "black swan" events—high-impact, low-probability threats that fall outside of historical models. Furthermore, it can struggle to quantify the positive benefits of security beyond loss avoidance, such as enhanced customer trust or enabling secure digital transformation initiatives that generate new revenue.

Future work in this field should focus on improving the data foundations for these models. This includes the development of industry-specific data lakes for cyber incidents to inform more accurate ARO estimates and standardized methodologies for valuing digital assets. The integration of these economic models into automated security platforms could also enable real-time, dynamic risk assessment, allowing organizations to adjust their security posture and investments based on a continuously updated view of the economic threat landscape.

5. Conclusion

This paper has established a comprehensive framework for evaluating cybersecurity investments through the rigorous application of economic metrics. The analysis, grounded in the hypothetical but realistic case of SecureFin, demonstrates unequivocally that cybersecurity initiatives can be transformed from intangible costs into quantifiable, value-driven investments. By leveraging established financial concepts such as Total Cost of Ownership (TCO), Annualized Loss Expectancy (ALE), Net Present Value (NPV), and Return on Security Investment (ROSI), we have provided a replicable methodology for justifying security expenditures in the language of business: risk reduction, return on capital, and value creation.

The key conclusion from this research is that a metric-driven approach fundamentally bridges the communication gap between technical security teams and executive leadership.

The case study results—a 275% ROI and a positive NPV of $848,177—present a compelling, data-driven narrative that surpasses fear-based appeals. This methodology allows Chief Information Security Officers to present cybersecurity not as a drain on resources, but as a strategic function that directly contributes to the organization's financial health and operational resilience. It enables informed prioritization, ensuring that limited resources are allocated to counter the threats that pose the most significant financial risk to the enterprise.

However, the efficacy of this model is entirely dependent on the accuracy and rigor of its inputs. The critical challenge remains the credible estimation of parameters, such as threat probability and potential loss. Therefore, the adoption of this framework must be accompanied by a commitment to enhancing data collection through cross-functional collaboration and the application of industry-standard risk quantification methodologies. Organizations must view this not as a one-time calculation, but as an ongoing process of refinement and validation.

Ultimately, the transition to an economics-based cybersecurity model represents a necessary maturation of the profession. It moves the discipline beyond technical controls and compliance checklists towards a sophisticated, business-aligned practice of risk management. By embedding these economic principles into their governance structures, organizations can make more rational investment decisions, optimize their security posture, and demonstrably prove that effective cybersecurity is not merely a defensive necessity but a definitive contributor to sustainable business growth and longevity.

References

1. ‘What Is Return on Investment (ROI) and How to Calculate It’, Investopedia. Accessed: Oct. 26, 2025. [Online]. Available: https://www.investopedia.com/terms/r/returnoninvestment.asp
2. ‘Net Present Value (NPV): What It Means and Steps to Calculate It’, Investopedia. Accessed: Oct. 26, 2025. [Online]. Available: https://www.investopedia.com/terms/n/npv.asp
3. ‘Total Cost of Ownership: How It’s Calculated With Example’, Investopedia. Accessed: Oct. 26, 2025. [Online]. Available: https://www.investopedia.com/terms/t/totalcostofownership.asp
4. ‘7 Costs of Maintaining Legacy Systems & How to Avoid Them’, DevSquad. Accessed: Oct. 26, 2025. [Online]. Available: https://devsquad.com/blog/costs-of-maintaining-legacy-systems
5. ‘What Is an ALE Formula? (And How To Use It)’, Indeed Career Guide. Accessed: Oct. 26, 2025. [Online]. Available: https://www.indeed.com/career-advice/career-development/ale-formula
6. ‘What Is ARO | Annualized Rate of Occurrence | Sangfor Glossary’. Accessed: Oct. 26, 2025. [Online]. Available: https://www.sangfor.com/glossary/cybersecurity/what-is-annualized-rate-occurrence
7. ‘Single Loss Expectancy - an overview | ScienceDirect Topics’. Accessed: Oct. 26, 2025. [Online]. Available: https://www.sciencedirect.com/topics/computer-science/single-loss-expectancy
8. KrishnaRahul, ‘Understanding the Total Cost of Ownership: What It Is and Why It Matters’, Cflow. Accessed: Oct. 26, 2025. [Online]. Available: https://www.cflowapps.com/total-cost-of-ownership/
9. ‘Risk Assumption - an overview | ScienceDirect Topics’. Accessed: Oct. 26, 2025. [Online]. Available: https://www.sciencedirect.com/topics/computer-science/risk-assumption
10. ‘What is an Advanced Persistent Threat (APT)? | CrowdStrike’, CrowdStrike.com. Accessed: Oct. 26, 2025. [Online]. Available: https://www.crowdstrike.com/en-us/cybersecurity-101/threat-intelligence/advanced-persistent-threat-apt/
11. ‘What is SIEM? | IBM. Accessed: Oct. 26, 2025. [Online]. Available: https://www.ibm.com/think/topics/siem
12. ‘Cost of Capital: What It Is & How to Calculate It’, Business Insights Blog. Accessed: Oct. 26, 2025. [Online]. Available: https://online.hbs.edu/blog/post/cost-of-capital

Utilizing Artificial Intelligence in Auditing and Investigations

Iavor Bachev

University of National and World Economy, Sofia, Bulgaria
ybashev@unwe.bg

Abstract. The integration of Artificial Intelligence (AI) into auditing and investigation processes represents one of the most transformative developments in the field of corporate governance and financial oversight. This paper explores the increasing utilization of AI in auditing, its benefits, challenges, and implications for the future of assurance and fraud detection by performing an empirical review of the literature and other relevant sources. Recent reports indicate that approximately 39% of internal audit functions currently utilize AI, with adoption expected to rise significantly by 2026. Through the review of existing academic and professional literature, this study identifies the key technologies being implemented—such as machine learning, robotic process automation, and natural language processing—and assesses their impact on audit efficiency, fraud detection accuracy, and overall audit quality. Furthermore, it highlights the emerging risks associated with AI adoption, including data quality issues, lack of technical expertise, algorithmic bias, and ethical or regulatory uncertainties. The paper also discusses the evolution of traditional fraud models toward AI-driven environments, introducing the concept of the "AI Fraud Diamond," which acknowledges the complexity and technical opacity associated with modern fraud schemes. The findings emphasize the need for robust governance frameworks, interdisciplinary collaboration, and the continuous development of auditor competencies to ensure that AI enhances rather than compromises the integrity and effectiveness of the auditing process. Based on the findings the AI is here to stay and its adoption will become wider and wider auditors and investigators will have to learn to adapt to the new dynamics set by AI.

Keywords. Audit, artificial intelligence, fraud detection.

1. Introduction

The Artificial Intelligence boom is an example of a technology breakthrough similar to the invention of the computer. Businesses have begun relying on AI more and more in their day-to-day operations. Although artificial intelligence is a relatively new phenomena in auditing a report by Protiviti [1] shows that about one in five finance organizations have achieved benefits related to cost and efficiency by utilizing AI. Moreover, AI can be used in most organizational units. Based on a recent Pulse report by the Institute of Internal Auditors [2] almost half of the internal audit functions have reported using some sort of AI in their processes, with 34% of the respondents currently prospecting future use of AI in their activities. A recent survey by Wolters Kluwer [3] indicates that 39% of internal auditors are currently utilizing AI in some capacity with an estimate of 41% intending to adopt AI in 2026. As with all new technologies the adoption of AI comes with some emerging risks. According to the 12th Annual Global Internal Audit Perspectives on Top Technology Risks survey [4] 59% of the respondents indicate the advanced AI systems as a significant risk in the next couple of years. Highest risks include but are not limited to security risks (hacking, data poisoning), privacy risks (rights & law violations) and operations risks such as systems errors or unexpected downtime. Authors such as Bankova [5] have foreseen and pointed out the

needed shift from traditional IT audit to cyber audit due to the increasing dependence on IT technologies.

2. Artificial intelligence in auditing & investigation process

Suyono et al. [6] provide an overarching review of how artificial intelligence can be integrated in the auditing process as well es outline some of the limitation of utilizing such in auditing. The authors have leveraged several major academic databases including but not limited to Scopus and Google Scholar, to screen a set of articles which can be selected for further analysis. Based on their findings auditors have used the following technologies – Machine learning (suitable for pattern recognition and fraud detection), Robotic Process Automation (suitable for data extraction, transaction testing) and Natural Language Processing (suitable for analyzing sets of unstructured data for outliers, red flags and other irregularities).

The authors further outline several challenges if using AI including:

- Data of insufficiently good quality which decreases the accuracy and effectiveness of the results provided.
- Lack of competency with AI, which impede effective adoption
- Issue related to ethics, transparency and accountability in situations where concerns related to fairness, bias and accountability arise.
- Expected regulatory intervention and uncertainty due to expected standards and regulations around AI usage and audit evidence etc.

The authors stress the need of accepting guidelines and frameworks on the use of AI in the audit process as well as the need for further research on the results of AI implementation in those processes such as the effect on audit quality, fraud detection rates, false positives etc). It has to be noted that such guidelines and frameworks could start to be used be used in preventing other risks to the organizations such as cyberfraud. As Bankova [7] states "hackers are constantly developing their skills and are increasingly taking advantage of the fear, insecurity and nativity of the average user". Such actors are likely to utilize AI to success in their goals as well making the detection of fraud harder for organizations.

Aitkazinov [8] explored the question of how Artificial intelligence tools and technologies can be integrated into auditing processes and increase their effectiveness and accuracy. Several areas are identified which can benefit from AI: automation of routine tasks, improved anomaly detection, real-time monitoring on processes and improved audit quality and risk assessment. The challenges identified by the author encompass several areas emphasizing the need of technical expertise and resources, algorithmic bias and the changing role of the auditor shifting to a more oversight-based activities requiring different mindset and skills. The author reaches the conclusion that businesses should invest in training the necessary staff in the best way to utilize AI in order to increase the AI literacy as well as stresses the need of collaboration between different parties in and outside organizations such as auditors, data scientists and oversight bodies to adopt frameworks and standards that ensure AI usage aligns with professional standards, judgement and ethical considerations. The author believes that the integration of AI tools has the potential to bring transformation to the audit practice due to the fact that it allows auditors to overcame traditional sampling methods (moving from sampling to full data analysis), improve the analysis performed and to reduce the time needed for mundane tasks.

Fedyk, A., Hodson, J., Khimich, N. et al [9] have analyzed the impact of Artificial intelligence and the impact on the quality of the audit. The results indicate that AI investment improves the quality with a 5% reduction in the likelihood of an audit misstatement and reduce audit fees (up to 0.9%). The authors also provide an overview of the employees that utilize AI

in the auditing sector- most of them being a young male with the majority being with technical background such as computer science or mathematics – most likely the skills needed to configure and use the AI tools properly. Last but not least is a significant rise in the AI talent in the recent years which signals higher and higher adoption of artificial intelligence utilization. Rafael Romero-Carazas at al [10] have conducted a literature review and bibliometric analysis in Scopus indicating that forensic auditing with the help of artificial intelligence have advanced enabling auditors to identify fraud faster and at an early stage and minimize the negative impact of fraud should it occur. The importance of identifying critical data is stated as well as the need of auditors to adapt to the advances in technology. Supriadi [11] has specifically examined the factor of Artificial Intelligence in detecting fraud in accounting audits. The conclusions of the author indicate and support the general consensus that AI can indeed improve not only the efficiency but the accuracy in detecting fraud. One of the main challenges posted by the author is the barriers of acceptance of AI in the accounting industry.

Fraud, investigation models and methodologies have been impacted as well by Artificial Intelligence. Traditional models like the fraud triangle have evolved. Zweers, B et al [12]. Have acknowledged that due to the technical advancements fraud has increasingly been depended on IT/AI systems and introduced the so called "AI fraud Diamond" which further develops the fraud triangle by adding a fourth element – technical opacity. The authors have used the traditional elements of Pressure (individual or organizational), Opportunity (process gaps and insufficient controls) and Rationalization (what is accepted internally) and shown how those manifest in environment where artificial intelligence exists. The fourth element – technical opacity- occurs when the environment is so complex and hard to understand that auditors and investigators cannot fully understand what is the situation. A variety of factors can lead to the technical opacity including but not limited to: constrained access, or even not enough knowledge. The authors further distinguish several types of fraud related to AI which can be used as a basis for auditors and investigators those are: input data manipulation (source data alteration), model exploitation and evasion, algorithmic decision manipulation (algorithms predicant an outcome/results based on bias or some other factor), synthetic fraud (fake identities or information facilitated by deepfakes) and AI ethics fraud (transferring the decision making and responsibility to AI for mismanagement or bad decisions).

A logical last question would be on how to audit and provide assurance that the artificial intelligence Ojewale et al. [13] have argued that although there are many tools that exists currently for AI auditing those can appear to be inadequate due to the inability of the users to access the data governing the behavior of the AI systems and apply consistent and holistic standards to provide integrity of the audit.

3. Conclusion

Artificial Intelligence has begun to redefine the auditing and investigative landscape, transforming how organizations detect irregularities, manage risks, and ensure compliance. The literature and empirical evidence reviewed indicate that AI not only improves audit efficiency and reduces costs but also enhances the accuracy of fraud detection and the quality of audit outcomes. Technologies such as machine learning, robotic process automation, and natural language processing enable auditors to move beyond traditional sampling methods toward comprehensive, data-driven analysis.

However, the rapid adoption of AI introduces new layers of complexity—ethical, technical, and operational—that demand vigilant oversight and new skillsets within the auditing profession. Risks related to data integrity, algorithmic bias, privacy, and regulatory

uncertainty underscore the importance of establishing clear standards and ethical frameworks governing AI use. Moreover, the evolution of fraud typologies, as illustrated by the AI Fraud Diamond, reflects the growing sophistication of digital deception, requiring auditors and investigators to adapt their methodologies accordingly.

Ultimately, successful integration of AI into auditing depends on a balanced approach—leveraging its analytical power while maintaining human judgment, ethical rigor, and accountability. Continuous learning, interprofessional collaboration, and proactive governance are essential to harness the full potential of AI as a force for transparency, accuracy, and trust in the audit and investigation processes. More research will be needed on how the landscape evolves in the near future. The rate of change should be proportional to the amount invested and wider adoption of the Artificial intelligence in more and more process.

References

1. Protivity., Harnessing the future: Protiviti's research on AI adoption, 2024. Available at: https://www.protiviti.com/sites/default/files/2024-11/research-brief-harnessing-future-research-ai-protiviti-2024.pdf
2. Internal Audit Foundation, 2024 North American Pulse of Internal Audit Benchmarks for Internal Audit Leaders. 2024. Available at: https://www.theiia.org/globalassets/site/resources/research-and-reports/pulse-of-internal-audit/2024-iia-pulse-report.pdf
3. Wolters Kluwer, Survey from Wolters Kluwer, 2025. Available at: https://www.wolterskluwer.com/en/news/new-survey-wolters-kluwer-internal-auditors-double-ai-adoption-2026
4. Protivity, From AI to Cyber – deconstructing a complex technology risk landscape: Assessing the results of the 12th Annual Global Internal Audit Perspectives on Top Technology Risks Survey from Protivity and the Institute of Internal auditors, 2024. Available at: https://www.protiviti.com/sites/default/files/2024-09/12th-annual-global-internal-audit-top-technology-risks-survey-iia-protiviti.pdf
5. Bankova, D., 2020. ABOUT CYBERSECURITY IN THE COMPANIES, IDES issue 1. Available at: https://www.ides.bg/media/1741/01-2020-diyana-bankova-1.pdf
6. Suyono, Windy Permata, Eka Septariana Puspa, Surya Anugrah, and Rio Firnanda. 2025. "Artificial Intelligence in Auditing: A Systematic Review of Tools, Applications, and Challenges". RIGGS: Journal of Artificial Intelligence and Digital Business 4 (2):3393-3401. Available at: https://doi.org/10.31004/riggs.v4i2.1024.
7. Bankova,. D, 2021. About the challenges of Cybercrime in the digital age. Yearbook of the UNWE, ISSN: 1312-5486; ISSN (online): 2534-8949; Available at: https://doi.org/10.37075/YB.2021.2.08
8. Aitkazinov, Arystanbek. 2023. "The Role of Artificial Intelligence in Auditing: Opportunities and Challenges". International Journal of Research in Engineering, Science and Management 6 (6): 117-119. Available at: https://journal.ijresm.com/index.php/ijresm/article/view/2740.
9. Fedyk, A., Hodson, J., Khimich, N. et al.,2022. Is artificial intelligence improving the audit process?. Rev Account Stud 27, 938–985 Available at: https://doi.org/10.1007/s11142-022-09697-x
10. Romero-Carazas R., Espíritu-Martínez A. P., Aguilar-Cuevas M., et al.2024, "Forensic auditing and the use of artificial intelligence: A bibliometric analysis and systematic review in Scopus between 2000 and 2024." Heritage & Sustainable Development Vol 6 No. pp 415-428. Available at: https://hsd.ardascience.com/index.php/journal/article/download/626/144/1933
11. Iman Supriadi. 2024. "The Audit Revolution: Integrating Artificial Intelligence in Detecting Accounting Fraud". Akuntansi Dan Teknologi Informasi 17 (1), 48-61. Available at: https://doi.org/10.24123/jati.v17i1.6279
12. Zweers, B., Dey, D. and Bhaumik, D., 2025. The AI-Fraud Diamond: A Novel Lens for Auditing Algorithmic Deception. arXiv preprint arXiv:2508.13984. Available at: https://arxiv.org/pdf/2508.13984
13. V. Ojewale et al.,2024 Towards AI Accountability Infrastructure: Gaps and Opportunities in AI Audit Tooling." arXiv preprint. Available at: https://arxiv.org/abs/2402.17861

Cybersecurity Risk Management and Threat-Led Penetration Testing for Financial Stability

Maya Zhelyazkova

University of National and World Economy, Sofia, Bulgaria
mayazh2@gmail.com

Abstract. Since the Covid-19 crisis and the following period in 2020-2022, the digital world has become standard in order to facilitate all spheres of human activity. The virtual environment leads also to saving resources. As a result, information technologies have evolved at an immediate pace and the financial sector, along with many other areas of activity, has increasingly begun to fuse into the digital space. The financial ecosystem evolved. However, in addition to this development, cyber threats in the financial operations and financial entities have also augmented, leading to the need to achieve a resilient and secure digital environment. This need has become a critical element for the functioning of the financial sector, even more so in terms of the interconnectedness of the financial system operators. This article focuses on the latest developments in relation to the measures to secure the proper functioning of the financial sector in the digital world and particularly on the threat led penetration testing designed to identify weaknesses and prescribe preventive measures for better operational resilience of the financial entities.

Keywords: Cybersecurity, cyber-attacks, cyber threats, threat-led penetration testing

1. Introduction

The world is changing at an incredible speed and the financial stability is of primary importance. In the past, we have evidenced that the consequences of the financial crises could have a devastating effect on the regular citizens' lives. The models nowadays for stronger financial Europe are more than ever repetitive. In the 2025 State of the Union address by the European Commission President von der Leyen during a plenary session in the European Parliament in September 2025, an annual speech that highlights the priorities of the work of the European Commission in the year to come, it becomes obvious that Europe is now developing in a new way. Many actions were proposed to support digital projects and transformations. The President indicated that "…we will massively invest in digital and clean tech" [1] and this way she revealed the forthcoming preparations for the digital euro, support for the European AI Gigafactories and other innovative ways to facilitate businesses and consumers for competitive (and independent) Europe.

In an international situation characterised by uncertainty, strong economic competitiveness and increasing levels of conflict, ensuring a secure virtual operating environment is essential for economic stability, especially in view of the interconnectivity and interdependence of the services. The activities of cybercriminals pose new hybrid risks to the economic stability. We are witnessing sabotage of undersea telecommunication cables, system disruptions in large banks and third-party suppliers, as well as increasing complexity of cyber-attacks and hacker skills. Joining up cyber protection efforts is essential since there is a growing need for international cybersecurity drivers. This should be accomplished at European/international and not only at national level.

Further to the development of the digital world, the international, European and national authorities focused attention on the risks of information and communication technologies (ICT). The recent trends in preserving cybersecurity include preventive measures to minimize the incidents, which could be increasingly influencing and modifying the financial sector. As part of the measures, in principle every 3 years threat lead penetration testing must be accomplished for the critical information technologies providers in the financial sphere. This is the exact focus of the current paper aiming to present the last tendencies in managing the cybersecurity aspects of the financial world.

2. The importance of addressing ICT risks

Nowadays, cyber threats are experienced daily and therefore the companies have to be resilient continuously. Recent statistics show that the world is facing over 2,200 cyberattacks per day. In the latest publication of the European Union Agency for Cybersecurity (eNISA) [2], it is reported that intrusion via ransomware is significant, phishing has become even an industry, and the AI has now its influence in the area. The number of high impact incidents has decreased whereas the diversified and continuous pressure targets to erode the resilience of the organisations.

Currently, the financial organisations are trying to find the balance between convenience (providing user friendly services to consumers), competition (several financial entities competing in the market, offering innovative services at low cost) and resilience (the ability of a financial entities/the financial sector as general to continue to carry out its tasks by anticipating and adapting to the threats and other relevant changes in the environment and by recovering from incidents). Obviously, this is one of the main managerial tasks nowadays.

Nowadays the number of high impact incidents decreases. Some of them are worth mentioning, however. Since the systems of the economic operators are entirely in digital format and this way they are also interconnected, an incident in one of the operators could have many implications for others. It becomes obvious that prevention measures must be in place to minimise and, as far as possible, avoid domino effect upon incident.

One clear example of massive implication for other entities is the case with Amazon Web Services (AWS) cloud services outage on 20 October 2025. AWS is a company providing internet-based cloud service connecting businesses and people using their applications or online platforms. Businesses using AWS pay a subscription fee to enable reliable online communication between their companies and customers. Many companies prefer to use the AWS services and depend on them to operate, rather than building and maintaining those services internally, which is far more costly and, in many instances, less secure. AWS has demonstrated computing power to handle the massive amounts of traffic that has come to its site ever since it started operating in 2006.

According to an HG Insights' report published in 2025 [3], AWS has a market share of 30% of the global cloud infrastructure market with 4.19 million customers geographically situated in North and South America, Europe and Middle East and Asia-Pacific region – this easily explains the significant global impact upon AWS outage. The incident in the cloud provider infrastructure affected banks (HALIFAX bank, LLOYDS bank), airlines (Delta Air Lines, United Airlines), media, many websites and applications. A number of online services including Venmo, Snapchat, Fortnite, Facebook, Prime Video and Amazon's cloud computing were not functional. CNN reported that Coinbase, a cryptocurrency exchange as well as the artificial intelligence company Perplexity had technical disruptions. Many customers were not able to view/access their funds.

The main cause of the incident seemed to be an issue in the Domain Name System, an internet location engine, converting user-friendly web addresses into IP addresses (a series of

numbers that other websites and applications can understand). The company kept its customers informed about the recovery steps undertaken. However, the situation proved the need for better preparedness and showed the importance of addressing the concentrated risk. The main aim should be to minimize and if possible, avoid the enormous effect in view of the interconnectedness of the operators – in the AWS case, the impact was from consumer applications to financial and public sector services.

Another interesting example dates back in July 2024 [4], when a software glitch during an update made by the US-based cybersecurity company CrowdStrike wreaked havoc across sectors globally. Airlines were forced to cancel flights, hospital disruptions were observed and businesses were not able to access their internal platforms. The event showed the fragility of the global economy in view of ICT service and the vastly spread dependence on computer systems.

At that time, numerous companies used CrowdStrike's cybersecurity software to block cyber threats. Upon updating their cybersecurity software, computers running with Microsoft Windows crashed because of the faulty way a code update issued by CrowdStrike interacted with Windows. Although, the company had significant experience with more than a decade of being present at the market, this result lead to blocking the functioning of many operators and later the declared costs were more than $5 billion in direct losses. Currently, app 10 000 companies use CrowdStrike.

In this case, obviously, more diversity in the cyber security providers could minimise the impact. At that time however, the dominance of a number of companies in the anti-virus and threat-detection field created risks.

Meanwhile, initiatives to prevent market disturbances have already started at national and supranational levels. In 2016, directives were given to the EU member states to adopt national strategies on network and information systems [5]. This horizontal measure covered three types of financial entities – credit institutions, trading venues and central counterparts. The countries were able to identify operators of essential services but only for those three types of financial entities. The exact rule for this identification (called size-cap rule) is later established in the NIS 2 Directive [6].

Currently, there are approximately 22 000 financial entities operating in the European Union. Every incident occurred in one of those entities concerns also others interconnected and poses risks for the functioning of the businesses but also for the consumers. To address this and to facilitate the digital functioning in the financial sector, and consequently to enhance the operating of the single market, the legislators proposed preventive measures for stability.

The EU financial sector is regulated by the Single Rulebook and the governance is provided by the European System of Financial Supervision. Up till 2022, there have been legislative differences and separate approaches in tackling ICT risks by the national authorities. This could be a factor distorting the competition between the financial entities operating in different member states and consequently, could occur to the detriment of the market participants. In addition, there were gaps or overlaps in digital operational resilience testing, which is not cost effective and showed the need for more cooperation and alignment at least at European level both in preventing incidents but also in mitigating their impact once they have occurred. As a result, new legislation was proposed targeting the updating of the ICT risk requirements such that they are not covered only by capital requirements, as was the case in the past. The new legislation targeted also to establish equal rules for financial entities when assessing and addressing ICT risk.

This brought consistency and more confidence in the EU financial system. Additional costs were avoided by minimising the risk of ICT disruptions. Also, the legal certainty was increased, more simple rules were introduced and in cross-border cases, more convergence

was established - the legislators introduced the European Regulation 2022/2554 on digital operational resilience for the financial sector [7], which entered into force on 16 January 2023 (with the abbreviation DORA – Digital Operational Resilience Act).

3. Ensuring control

As mentioned, some financial entities were already part of the ecosystem of the NIS-2 Directive in 2022. The national competent authorities as well as the three European Supervisory Authorities (the European Banking Authority, the European Securities and Markets Authority and the European Insurance and Occupational Pensions Authority) participate in the workings of the Cooperation Group and the computer security incident response team CSIRTs, both established under the same Directive. This enabled the smooth exchange of information and lessons learned experience.

In order to address the need for more joint efforts to prevent the cyber-attacks and the related risks in the financial sphere at European level, DORA introduced new cybersecurity rules. Increased investment in security and regular risk testing is required to ensure the good functioning of the financial entities' businesses. DORA established a broad resilience framework in the financial sector for both financial entities and their ICT providers that are identified as critical.

Many financial institutions prefer to use providers for software solutions, cloud services, data processing. This has led to a risk of cyber threat to those providers, which may also compromise financial institutions themselves. With the introduction of DORA in the European Union, financial institutions must carry out thorough procedures to manage risks associated with external providers and carry out regular security tests. The European Banking Authority, the European Securities and Markets Authority and the European Insurance and Occupational Pensions Authority are the leading supervisors and ensure the monitoring of critical service providers in the EU on the risks they can pose to the financial sector. On-site and remote inspections are foreseen and recommendations are to be issued.

In 2023, the mentioned European Supervisory Authorities issued a report on the state of information and communication technology providers in the EU [8]. The document indicates that there are approximately 15000 service providers directly serving the EU financial sector. They support many critical or important functions of the financial entities. Separately, some of the providers consequently use sub-providers and thus the number of ICT providers reaches around 20 000. Of course, the risk of using these services depends on the level of criticality of the service and it is estimated that around 9 000 service providers serve activities for financial institutions identified as critical or important. Some of those services are development and deployment of software and applications, network infrastructure services (excluding telecommunication services), data centres (physical), ICT advice and managed ICT services, information security and cybersecurity services (including control and monitoring, intrusion testing, security operations centres, etc.), cloud computing and data analytics and other data related services. Most of the companies providing these services focus their work mainly on one area. Software and application development is the activity where most financial institutions use suppliers.

The critical ICT service providers were first identified by the ESAs in November 2025. The number of entities identified under the EU framework are currently 19 and they specialise in different services.

The financial entities themselves manage the ICT third-party risk while the member states competent authorities accomplish the financial supervision of ICT risks. The oversight function of the European Supervisory authorities complements the ICT risk supervision.

4. International practices

At the international stage, there are a number of frameworks for operational resilience established in the UK, the USA, Singapore, India, others. Globally, the G7 Cyber Expert group functions as a working group that meets annually to discuss the cybersecurity policies and strategies in the member countries of interest to the financial sector. The group was established in 2015 and the chairs are representatives of the Bank of England and the US Treasury. The countries (Canada, France, Germany, Italy, Japan, the UK and the USA) exchange information, cooperate and discuss incident responses during their four meetings per year. On annual basis, an incident response test is organised and every four years cross-border cyber exercises take place. The work is reported to the G7 Finance Ministers and Central Bank Governors. Non-enumerated members are the European Banking Authority, the European Central Bank and the European Commission.

One of the latest topic of discussion of the G7 Cyber Expert Group is AI and the risks it could pose to the financial system [9]. After the development of the ICT systems, AI is the next step of the expanding innovation that could further transform the financial system. It could be used by the financial entities and authorities to improve their operations, analyses and risk management. However, AI is explored also by malicious actors and particularly generative and agentic AI could lead to new cybersecurity risks. In this view, it seems inevitable for the authorities to establish partnerships with the academia and technology firms to counterfeit those threats.

In the constantly evolving environment nowadays, a number of cyber security and resilience frameworks were developed to become operational internationally. Some of them are listed below with the aim to generally present the cyber security measures globally.

NIST Cybersecurity Framework

The National Institute of Standards and Technology (NIST) within the US Department of Commerce has established the second version of its cybersecurity framework [10] and provides support for the entities to manage their cybersecurity risk. The system of cybersecurity outcomes is provided and no instructions are given how they have to be achieved, which is left for the managers of the risk. This framework is available further to multiyear cooperation between the industry, academia and the government of the USA and is considered to be one of the best frameworks. It has horizontal context and can be applied no matter the size and the sphere of work of the entities.

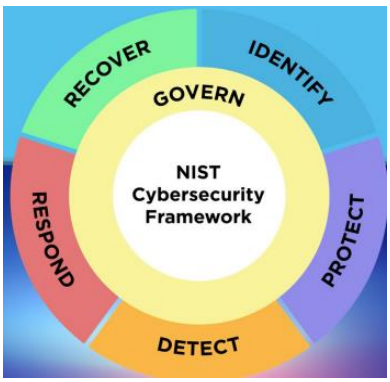


Figure 1. Illustration of NIST Cybersecurity Framework (Saurce: National Institute of Standards and Technology, US Department of Commerce)

The core components of the NIST Cybersecurity Framework are identify (understand the entities' environment), protect (ensure the delivery of critical services), detect (recognise cyber threats rapidly), respond (actions taken upon cybersecurity incident) and recover (restoring services weakened further to an incident).

The UK example

The UK as a member of the G7 Cyber Expert group has its own framework for improving the resilience of the financial sector. The country authorities highlight the importance of assessing the cyber risk by the individual companies and the financial markets infrastructures such that they are well prepared for and can adequately respond to cyber threats and incidents. The following tools are used for cyber assessment – CBEST, STAR-FS and CQUEST.

CBEST (Critical National Infrastructure Banking Supervision and Evaluation Testing) is a framework for regulating authorities in the UK to work with companies using a mock cyberattack on the people, processes and technology (in live production systems). The basis for the simulated attacks includes the current cyber threats trends with the aim to improve the company's resilience. In testing the firm's defenses, the regulators can assess their ability to detect and respond to cyber threats. The simulation itself is accomplished by accredited service providers and they act following legal, ethical and moral obstructions in view of the fact that the testing is accomplished on live production systems.

Every year, a CBEST thematic report is published to inform the private sector about the current cyber threats such that the entities could review and strengthen their ICT systems and tools. As regards 2025, the report is expected to include details in respect to the management of ICT and cyber risks. In the 2024 report [11], it is highlighted that there are still gaps in the foundational cyber defenses. The entities should introduce improvements as regards assessing the cybersecurity risk to assets and individuals, cyber risk management, detection of cyber threats and cyber incident response to mitigate impact. The firms still show weaknesses in their threat intelligence, namely on the integration of threat intelligence with business lines. In 2023, the need for a strong foundation of cyber hygiene is highlighted, i.e. practices for preventing cyber incidents like training and awareness raising campaigns, maintenance of assets and detection capability. CBEST help firms assess their resilience and review their capabilities against attackers looking to disrupt important business services.

STAR-FS (Simulated Targeted Attack and Response assessments for Financial Services) is a tool to assess the cyber resilience of the important services of firms. This way the firms and national authorities have better view of the vulnerabilities and can take remedial actions. Threat led penetration testing is encouraged, which mimics the action of cyber threat actors. Financial institutions interested in the STAR-FS assessment can establish contacts with their supervisory authority and gain more information about their technical capabilities to protect, detect and reply to cyberattacks (including systems, people and processes).

CQUEST (Cyber resilience questionnaire) is used to assess cyber risk and resilience capabilities mainly in the UK bank sector but also other firms. It contains a questionnaire with 50 questions covering 6 domains: Governance and Leadership, Identify, Protect, Detect, Respond and Recover. The supervisory authorities might provide guidance further to the information received from the questionnaire.

Cybersecurity framework in Singapore

Singapore's cybersecurity framework represents a centrally coordinated approach to protect the critical information infrastructure. The legal foundation is in their Cybersecurity Act, which was recently amended in 2024. The updated legal framework for oversight and maintenance of the national cybersecurity includes new categories of regulated entities. The

Cyber Security Agency can respond when treats materialise. Recognizing that cybersecurity is a shared responsibility, Singapore also established a structured framework for sharing information and thus strengthening awareness and improving coordinated responses. At the same time, a licensing framework for cybersecurity service providers has been implemented to raise professional standards. There is a strong emphasis on people to nurture a vibrant cybersecurity ecosystemand sustain a talent pipeline.

India

In India, the authorities recognise the importance of cyber risk in a high-risk category. They accomplish public consultations and organise impact analyses with the aim to periodically review the landscape and issue guidelines. In the banking sector, the authorities organise a specialised IT supervision as from 2015. They prepare cyber mapping exercise and request enforcement and corrective measures, if needed in review of 24 different security domains (such as network security, application security, database security). Incidents have to be reported to the reserve Bank of India within 2-6 hours. Security Operations centre are available non-stop. Dedicated agencies have been established to protect the digital infrastructure.

5. TIBER-EU framework and Thread-Led Penetration testing

In Europe, TIBER-EU framework (Threat Intelligence-Based Ethical Red Teaming) provides operational requirements and guidance on how to perform testing, to leverage the learning based on the test results and as a result to improve the cyber resilience of the financial entities in the EU member states. TIBER-EU tests mimic the tactics, techniques and procedures of real-live attackers. It is tailor made to simulate an attack on the critical functions of the entity as well as its people, procedures and technologies. The test will reveal the strengths and weaknesses of the financial entity and will provide information for updating its cybersecurity level.

Following similar initiatives in the UK (and the Netherlands also), the TIBER-EU framework was built jointly by the European Central Bank (ECB) and the national central banks in the EU. It was published in May 2018 and a new edition of the Guidance is available as of January 2025 [12].

The requirement to test the ICT environment of financial entities is established with DORA regulation and has to be performed annually. Some organisations are subject to an additional testing – Threat-Led Penetration Testing (TLPT), which is required for financial entities whose failure would have systemic impact.

The usual testing aims to strengthen the security level of the entities while the TLPT has to confirm that the security level of the entity is appropriate. For more information as regards both TIBER-EU and TLPT, please see Table 1.

Table 1. Indicators as regards TIBER-EU framework and DORA TLPT

Indicator	TIBER-EU	DORA TLPT
Legal basis	NIS 2 Directive	DORA
Date of publication	May 2018	Dec 2022
Adoption	Voluntary	Mandatory

Mandatory	Voluntary framework	Legally binding regulation
Countries of application	Austria, Belgium, Czech Republic, Denmark, Finland, France, Germany, Iceland, Ireland, Italy, Liechtenstein, Luxembourg, Malta, the Netherlands, Norway, Portugal, Romania, Slovakia, Spain and Sweden.	EU member states
Mandatory elements	Four areas and two types of requirements are mandatory	Does not fully cover all the requirements of TIBER – EU framework. Compulsory for the financial entities across EU at frequency chosen by the TLPT authority (generally, 3 years).
Use of internal testers	Not envisaged	Possible under certain conditions
Purple teaming (joint testing of the red team (testers) and the blue team (staff from the attacked institution)	Not mandatory	Mandatory in the closure phase, similarly to the replay workshop
Sectors approach	Sector-agnostic approach	No sector specific or entity specific requirements
Selection of experience testers and intelligence providers	TIBER-EU procurement guidelines	Requirements in Art 27 of DORA and Regulatory Technical Standards specifying elements related to thread led penetration testing

In brief, the TLPT is a large scale exercise that simulates a real cyber attack on the financial entity. The TLPT includes physical and virtual testing. The physical test checks the possibility for on-site penetration and the virtual one is addressed to the ICT resources. The TLPT also ensures the coverage on the financial entities' employees – who can be a target of social engineering or blackmailing. Main principle during the testing is to achieve success by any means.

The main difficulties during TLPT are connected to:

- High operational risk since the test is accomplished on live production environment; there is a risk of service disruption or data lost or modification;
- ICT maturity requirements – in case the financial entity does not have a sufficient maturity, it should not go under a TLPT since they will not be resilient to handle the testing;
- Complexity of coordination – the tests involve different participants and coordination could be challenging while maintaining secrecy and compliance.

- Confidentiality and access controls – information must be shared only on a need-to-know basis; the governance could be complicated;
- Expertise - the TLPT require high skilled testers and thread intelligence providers. Searching for such experts can be difficult, time consuming and costly;
- The active red teaming phase usually takes place for 12 weeks and requires good planning and resource management;
- Risk management burden – the financial entities which undergo a TLPT takes full responsibility for risk management and continuous monitoring during the test.

The TLPT is to be accomplished in accordance with the TIBER-EU framework in force as of 8 July 2025. The criteria used for identifying financial entities that have to perform TLPT are specific and defined in a dedicated Regulatory Technical Standards as regards the requirements to use the internal testers, the requirements in relation to scope, testing methodology and approach for each phase of the testing, results, closure and remediation stages and the type of supervisory and other relevant cooperation needed for the implementation of TLPT and for the facilitation of mutual recognition.

The financial entities that have to undertake a TLPT are identified as per the principle of proportionality. Two-layers approach is applied and the following factors are assessed: impact (what will be the extend of impact in case of disruption of the services of the financial entity); financial stability (systemic impact at EU and national level) and ICT risk (level of ICT). The entities that are functional in core financial services and play a systemic role are subject to TLPT. It is possible that a TLPT authority decides to opt-out or opt-in a financial entity. Also, a TLPT authority may decide to perform a pooled or joint TLPT.

Pooled testing and joint testing

The pooled testing is accomplished when several financial entities will be in the scope of the TLPT, for which an ICT third party services provider will procure an external tester but only if it is expected that the non-pooled test have an adverse impact on the quality or security of services to customers that are entities falling outside of the scope of DORA or the confidentiality of the data related to such services. There are specific requirements for the pooled testing in connection to the remediation plan, the cooperation of the TLPT authorities and the attestation.

The joint testing involves several financial entities using the same ICT intra group service provider or belonging to the same group and using the same ICT systems.

The TIBER-EU framework includes four areas and two types of requirements including some optional requirements. The adoption of the framework is voluntary, if the respective country authority decides to adhere to the framework but once adopted the authority shall implement mandatory requirements.

It is possible to have cases of pan-European competent authorities. For trade repositories, credit rating agencies or critical benchmark providers, the European Securities and Markets Authority is the competent authority. The ECB is competent for credit institutions classified as significant (in accordance with Art. 6(4) of Regulation EU N1024/2013).

Participants

In TLPT, we have the following teams involved in the process:

- TLPT cyber team – includes the responsible people in the respective TLPT authority;

• Control team or called also white team – consists of the responsible people within the financial entity who manage the ongoing TLPT exercise. It deals with procurement, risk assessment, risk management;

• Blue team – consists of the responsible people protecting the financial entity against cyber threads; they do not have information about the testing;

• Threat intelligence provider – responsible for mimicking hacker information gathering;

• Internal and external testers – similar to the red team of TIBER-EU framework; DORA permits the use of both internal and external testers.

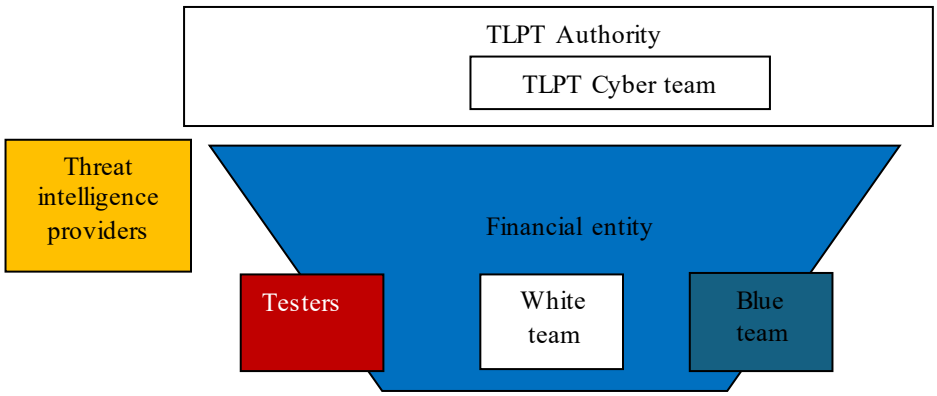


Figure 2. TLPT participants

Phases of TLPT

In general, every 3 years, threat lead penetration testing has to be accomplished for the critical ICT providers of services in the financial sphere. The responsibility lays within the financial entity – it accomplishes the test and bears all risks in relation to it. The process consists of three main phases – preparation, testing and closure.

During the ***preparation phase*** the control team is formed, the scoping formulation is identified and the testers are selected. In principle, this is the phase when good planning is to be accomplished. The TLPT authority should anticipate and inform financial entities about the forthcoming test and inform about the requirements. As next step the financial entities should start searching for suitable testers and threat intelligence providers as soon as possible after the notification and before the official start of the testing. Therefore, the preparation phase starts with the notification from the TLPT authority and includes procurement activities and risk management preparation. The phase is characterized with the drafting of the initiation documents (3 months) and documents to specify the scope of the testing (6 months).

In the ***testing phase***, we have the threat intelligence part with scenarios which have to be tested during the red teaming part and the active red teaming test has to be at least 12 weeks.

The ***closure phase***, which is similar to the one in the TIBER-EU framework, starts with the revealing of the TLPT to the blue team. The blue team and the red team replay some defensive and offensive actions carried out during the TLPT and a purple team exercise also takes place. The financial entity drafts a test summary report and a remediation plan and sends them to the TLPT authority. The process closes with an attestation from the TLPT authority.

Currently, it seems that there is a limited number of testers and threat intelligence providers in the market. Therefore, the requirements for the testers are not rigorous. In addition, there is a possibility that the providers do not cover the requirements but only in cases

when the financial entities mitigate any additional risk this introduce. Nevertheless, the threat-led penetration testing has established a good ground for better digital functional of the financial systemin the EU.

6. Conclusion

In conclusion, the dynamic evolution of the financial sectortowards digitalisation has led to new cybersecurity risk management practices and the adoption of new frameworks for cybersecurity and threat-led penetration testing. In this way, not only the regulators reinforce operational resilience but also foster greater alignment and collaboration across borders. As cyber threats continue to grow in sophistication and frequency, the financial sector must remain constantly ready, investing in regular testing, skilled personnel and adaptive frameworks. Eventually, the ongoing commitment to strengthening cybersecurity will underpin financial stability, ensuring that the interconnected global financial ecosystem can withstand and recover from disruptive incidents while maintaining trust and confidence among stakeholders.

References

1. The 2025 State of the Union Address by the European Commission President von der Leyen, Available at https://ec.europa.eu/commission/presscorner/detail/en/speech_25_2053
2. ENISA Threat Landscape 2025 publication in October 2025 - Available at ENISA Threat Landscape 2025_v1.2.pdf
3. Report "The AWS Ecosystem in 2025" by HG Insights, Available at https://explore.hginsights.com/aws-ecosystem-in-2025?utm_source=blog&utm_campaign=aws-ecosystem-in-2025&utm_medium=organic&utm_term=ug
4. https://edition.cnn.com/2024/07/19/tech/tech-outage-crowdstrike-software
5. Directive (EU) 2016/1148 concerning measures for a high common level of security of network and information systems across the Union (Available at https://eur-lex.europa.eu/eli/dir/2016/1148/oj/eng
6. Directive (EU) 2022/2555 on measures for a high common level of cybersecurity across the Union, (NIS 2 Directive) (Available at https://eur-lex.europa.eu/eli/dir/2022/2555/oj/eng
7. Regulation (EU) 2022/2554 of the European Parliament and of the Council of 14 December 2022 on digital operational resilience for the financial sector and amending Regulations (EC) No 1060/2009, (EU) No 648/2012, (EU) No 600/2014, (EU) No 909/2014 and (EU) 2016/1011, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32022R2554&qid= 1759041525129
8. ESA 2023 22 - ESAs Report on the landscape of ICT third-party providers in the EU (Available at https://www.esma.europa.eu/sites/default/files/2023-09/ESA_2023_22_-_ESAs_report_on_the_landscape_of_ICT_TPPs.pdf
9. G7 Cyber Expert Group Statement on Artificial Intelligence and Cybersecurity, September 2025 (Available at https://home.treasury.gov/system/files/136/G7-Cyber-Expert-Group-Statement-AI-and-Cybersecurity-2025.pdf
10. The NIST Cybersecurity Framework (CSF) 2.0 – Available at the following link https://nvlpubs.nist.gov/nistpubs/CSWP/NIST.CSWP.29.pdf
11. 2024 CBest Thematic report; Available at the following link https://www.bankofengland.co.uk/financial-stability/operational-resilience-of-the-financial-sector/2024-cbest-thematic
12. TIBER-EU Guidance for Service Provider Procurement (link https://www.ecb.europa.eu/pub/pdf/annex/ecb.tiber_eu_service_provider_procurement_2025.en.pdf?1d229f2191835b83770d593a44f69b14)

Countering Generational Disenfranchisement: Bipartisan Strategies to Enhance Homeownership for Millennials and Generation Z

Boyan Mitrakiev

University of National and World Economy, Sofia, Bulgaria
bnikolaev@unwe.bg

Abstract. The paper reviews current trends in US youth home ownership and links the declining rates across the Millennials and Genz population to potential disenfranchisement and radicalization as evidenced by existing empirical research and theory. As American democracy finds itself in a precarious position characterized by political violence, hate speech and policy polarization, the authors suggest that building an across the isle political consensus for improving home ownership among the youngest generations can be used as a tool for moderation and normalization. A brief review of a cross-ideological imperative on the topic proves that there is a realistic and achievable political consensus to be reached on policies that sustainably improve home affordability without imposing significant political costs on either major party even in the current political and media climate.

Keywords. Home ownership, American democracy, millennials, genz, radicalization, liberalism, conservatism, Judeo-Christian values.

1. Introduction

Home ownership is the essence of the American dream, symbolizing stability, predictable wealth accumulation, an anchor for civic engagement and local community integration, a safe space to raise a family and ultimately a valuable asset to pass down to the future generations, serving as a minimum guarantee that they will enjoy a better life and pursue higher opportunities than their parents.

However, recent trends indicate a sharp decline in homeownership rates among Millennials (born 1981–1996) and Generation Z (born 1997–2012), compared to the Silent Generation, Baby Boomers and Generation X at similar ages (Gregory, 2023; Gregory & Bloodworth, 2023). This shift is driven by factors such as wage stagnation, rising housing costs, student debt, and delayed life milestones like marriage and children (Isom, 2024; Gregory & Bloodworth, 2023).

While these barriers exacerbate economic insecurity, their broader implications extend to the sustainability of democracy and capitalism. Declining homeownership intensifies wealth inequality, fosters social re-stratification, and breeds resentment toward market systems, potentially leading to democratic apathy, radicalization, and a shift toward socialist ideologies (Forrest & Hirayama, 2018; Franc & Pavlović, 2018; Faria, 2022).

This paper argues that without intervention, these trends could undermine the mutual reinforcement of property ownership, democratic participation, and capitalist prosperity, as envisioned in the concept of a "property-owning democracy" (Vitali, 2023).

2. The threat of declining home ownership among millennials and generation z to democracy and capitalism in the United States

In her study, Gregory (2023) finds significant gaps in both median income and home ownership between the Silent Generation, Baby Boomers and Generation X on one hand and the Millennials and Generation Z, on the other. According to her, at the age of 30, "the median non-college-educated Millennial made 23% less than the median Silent Gen'er, 12% less than the median Boomer, and 4% less than the median Gen Xer" (Gregory, 2023). This gap in median income is an underlying foundation for a similar gap in home ownership, which however is exacerbated by the general inflationary processes for homes in the recent decades. The gap in some of the cohorts is especially striking and provides an obvious factual underpinning for the anecdotal evidence of intergenerational tensions today. For instance, at age 30, "non-college-educated Boomers were about twice as likely as non-college-educated Millennials to own homes: 49% vs. 24%, respectively" (Gregory, 2023) (Figure1.):

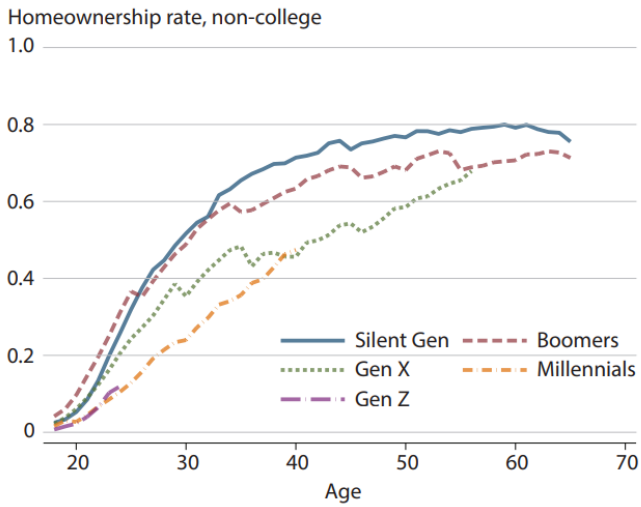


Figure 1.

The income gap among college educated cohorts closes sharply with Millennials at the age of 35 actually earning more than their Gen X and Boomer counterparts (Gregory, 2023). This phenomenon can be explained to some extent by the rising returns of college education across generations as argued by Autor (2024), though the financialization of the economy and the shift towards services-based entrepreneurship based on the coastal rural agglomerations may also be playing a part.

Yet, the home ownership gap in the college educated cohorts is also present as "at age 30, 38% of college-educated Millennials owned homes, less than the 54% of Boomer college-educated workers and even the 49% of Boomer non-college-educated workers of the same age" (Gregory, 2023) (Figure 2.):

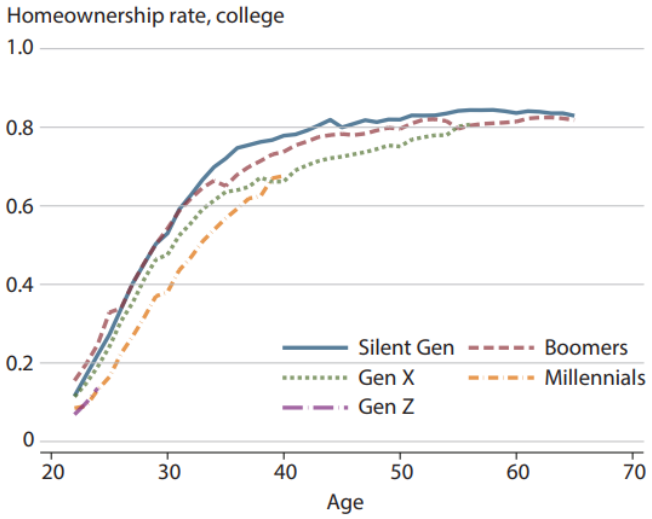


Figure 2.

Gregory (2023) argues that the persistence of the home ownership gap across generational groups who do not display an income gap point to "differing incomes and labor market opportunities are only part of the story". Inflationary pressures and wealth inequality are almost certainly putting negative pressure on the home ownership of younger generations today. A steeper entry level to the home market results in nearly half of young people between 18 and 29 now living with their parents – the highest share since the Great Depression (Isom, 2024) and almost doubling the same number during the post-WW2 economic boom. Isom (2024) also points to the prevalent lifestyle decision to delay marriage today as another barrier to owning a home by successfully paying down a mortgage as the AI Housing Center has found out that having a second borrower (most often a spouse) is significantly reducing the risk of default.

As already pointed out, the general financialization and pro-inflationary shift of the US economy has also had a profound impact on housing affordability and worker compensation. The Economic Policy Institute (2021) has calculated that the average annual growth in productivity and compensation for the period of 1948 – 1979 has been 2.5% and 2.1% respectively with the two diverging afterwards and the respective number being 1.4% and 0.6% for 1979 – 2025 (Figure 3):

Stagnating compensation levels have led to the median age of a first-time home buyer increasing to 38 years while the typical repeat buyer increasing further to 61 years with the median age of all buyers peaking at 56 years old (NAR, 2024). The numbers have increased remarkably throughout the last decades – for instance, in 1981, the median age of a first-time buyer was 29, the median age of a repeat buyer was 36 and the median age of all buyers was 31 (NAR, 2024).

This disparity is bound to create tensions between generations which have recently started to emerge from anecdotal evidence (Gaviola, 2019) to empirically established observations (Pisano, 2025; Bass, 2025) backed by statistical data that establishes that the biggest beneficiaries of the substantial run-up in home prices have been pre-late 1990s buyers (Macdonald, 2015).

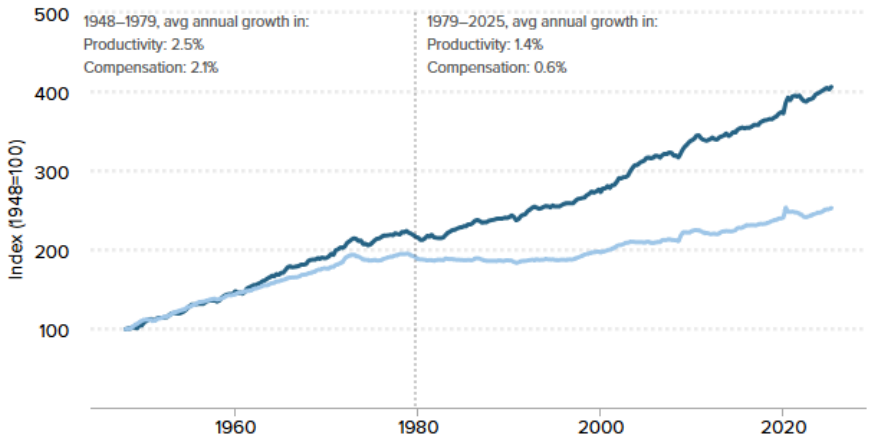


Figure 3.

A nationally representative poll conducted by Clever Offers, found out that American generations have started pointing the finger at each other regarding who is to blame for the housing affordability crisis that has taken over the country (Pisano, 2025) (Figure3.):

Although all generations seem to recognize the problems in the housing market, who's responsible is a more open question. Around a third of all respondents (31%) blame millennials most for the current affordable housing crisis, just ahead of boomers (27%). Notably, 35% of millennials selected boomers, while only 25% of boomers picked themselves. Meanwhile, 33% of boomers blamed millennials, while only 21% of millennials pointed the finger at their own generation. Somewhat surprisingly, respondents also see millennials (30%) as most responsible for the lack of housing inventory, as opposed to boomers (24%), who largely led the policy, planning, and business decisions that shaped the current market over the previous decades. Again, just 19% of millennials see themselves as most responsible for lack of housing inventory, while 30% of boomers feel this way. Meanwhile, only 24% of boomers chose their own generation, compared to 32% of millennials who selected boomers.

The same poll also found out a startling cleavage between the general perceptions of boomers and genz as home owners. The boomers are almost universally perceived as being most financial responsible (66%), most likely to oppose affordable housing (42%), least likely to buy a home as a status symbol (18%) but also the most likely to report neighbors to the HOA while not following neighborhood policies themselves (37%) (Pisano, 2025). On the other hand, genz are perceived by far as the least financially responsible (57%), the most unrealistic about what they deserve in a home (57%), most likely to overpay for a home out of desperation (40%), most entitled when it comes to home ownership expectations (31%) but also the biggest victim of the current housing crisis (45) (Pisano, 2025).

Higher interest rates and longer mortgage terms will exacerbate the perceived plight by millennials and generation z as they will be the first generations to face higher total projected mortgage repayments during the second half of the mortgage (Hamptons, 2024). This will squeeze this generation at a time when they are starting families or pursuing significant advancements in career and income, expecting these to translate into pleasure or substantial savings and investments. Researchers from the Grattan Institute in Australia have found out that millennials will also have higher expected total mortgage repayments as a share of expected total income during the loan, almost doubling the number for baby boomers (Khadem, 2023; Moloney and Coates, 2022; Coates, Moloney and Clay, 2022).

Finally, it is important to note that the general affordability crisis may create distorted perceptions among the young generations who are at the early stages of acquiring homes and repaying the mortgage as a simply temporal factor. While the general affordability hurts voters of all ages, the fact that it takes place at a time when millennials are still struggling with the first years of repaying the mortgage and the generation z are entering the labor force boosts their subjective vulnerability to the problem. The National association of home builders (2025) publish an annual "graph" of home affordability that indicates the scale of the affordability problem's impact as nearly 60% of US households (76.4 million out of a total of 134.3 million households) are unable to afford a $300 000 home with the median-priced new home in 2025 being priced at $459 826:

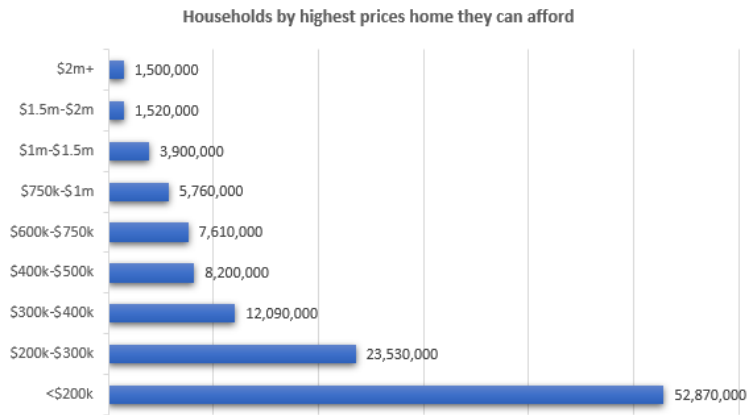


Figure 4.

The National association of home builder's (2025) assumption is actually underestimating the burden that the millennials and generation z are taking to finance home ownership as it is based on conventional underwriting standards that assume the cost of a mortgage, property taxes and property insurance not exceeding 28% of household income (both generations are observed to taking more than 30%). In other words, the minimum income needed to finance a purchase of a $200 000 home with a mortgage rate of 6.5% is $61 487, meaning that at the base of the graph there are 52.87 million American households unable to afford even this home.

3. The foundations of home ownership in America's Democratic and Capitalist system – from Greek philosophy to Judeo-Christian values

Philosophers such as Aristotle, John Locke, and Alexis de Tocqueville laid intellectual groundwork emphasizing property's role in balanced governance, while American founders like Thomas Jefferson and James Madison adapted these ideas to envision a republic of independent proprietors. Aristotle, in Politics, argued that moderate property ownership by a robust middle class was essential for political stability, preventing the extremes of oligarchy and mob rule. He posited that "the most perfect political community must be amongst those in the middle rank," as they "neither covet... nor are coveted by others," fostering a balanced polity where virtue prevails over factionalism (Aristotle, trans. 2009). For Aristotle, excessive wealth bred arrogance, while poverty incited sedition; thus, "it is best to have property private,

but to make the use of it common," ensuring subsistence without communal inefficiencies (Aristotle, trans. 2009). This middle-class emphasis stabilized governance by mediating class conflicts, a principle echoed in America's capitalist ethos where homeownership democratizes wealth accumulation.

John Locke extended this in Second Treatise of Government, viewing property rights as natural and pre-political, derived from labor: "every man has a property in his own person: this no body has any right to but himself. The labour of his body... are properly his" (Locke, 1689). Property encompassed "lives, liberties and estates," essential for liberty against arbitrary power (Locke, 1689). Government existed to protect these, as "the great and chief end... of men's uniting into commonwealths... is the preservation of their property" (Locke, 1689). Locke's labor theory justified capitalist acquisition while limiting it to prevent spoilage, aligning with democratic consent: Taxes required majority approval, safeguarding individual estates (Locke, 1689). This framework influenced America's Bill of Rights, embedding property as a bulwark against tyranny.

Alexis de Tocqueville, observing 1830s America, praised equality of conditions—facilitated by partible inheritance—as stabilizing democracy. In Democracy in America, he noted that "land is the basis of an aristocracy," but egalitarian division created smallholders, reducing class strife: "No sooner was the law of primogeniture abolished than fortunes began to diminish" (de Tocqueville, 1835). This fostered civic engagement, as "everyone has property of his own to defend" (de Tocqueville, 1835). Yet, he warned of risks: Excessive equality could breed individualism, but property ownership countered this by tying citizens to community interests, ensuring democratic vitality (de Tocqueville, 1835).

These ideas informed American founders. Thomas Jefferson idealized yeoman farmers as "the most valuable citizens," embodying republican virtue through self-sufficiency: "Cultivators of the earth are the most valuable citizens" (Jefferson, as cited in Phillips, 2018). Property ownership ensured independence, vital for democracy: "The small land holders are the most precious part of a state" (Jefferson, as cited in American Heritage, n.d.). James Madison, in Federalist No. 10, linked property to republicanism: "From the protection of different and unequal faculties of acquiring property... results" factional diversity, managed through extended republics (Madison, 1787). Government protected "property of every sort," including rights, to sustain liberty (Madison, 1792).

Jeffers and Adams drew their vision on integrating property into the republic's framework from Judeo-Christian tradition. Encompassing the Hebrew Bible (Tanakh/Torah for Judaism) and the Christian Bible (Old and New Testaments), it has profoundly influenced American political thought, embedding property ownership with moral, stewardship, and communal dimensions.

In Judaism, the Torah (first five books of the Hebrew Bible) frames property ownership as a sacred trust from God, emphasizing inheritance, justice, and limitations on alienation. This philosophy underscores homeownership's importance in fostering communal stability, a precursor to democratic equity. Central is the concept of nachalah (inheritance), where land is apportioned as a perpetual family holding, reflecting God's covenant with Israel. For instance, in Numbers 26-27 and 36, the Torah addresses inheritance rights, notably for Zelophehad's daughters, mandating: "The inheritance of the people of Israel shall not be transferred from one tribe to another, for every one of the people of Israel shall hold on to the inheritance of the tribe of his fathers" (Numbers 36:7, ESV). This ensures equitable distribution, preventing concentration of wealth and mirroring capitalist ideals of opportunity while curbing excesses through divine oversight.

Leviticus 25:23 explicitly limits absolute ownership: "The land shall not be sold in perpetuity, for the land is mine. For you are strangers and sojourners with me" (ESV). Here,

humans are stewards (gerim v'toshavim), not owners, aligning with sabbatical and Jubilee years where debts are forgiven and lands revert (Leviticus 25:10-13). This cyclical restoration combats inequality, philosophically underpinning democratic safeguards against oligarchy. Deuteronomy reinforces this with inheritance laws: "If a man has two wives... he may not treat the son of the loved as the firstborn in preference to the son of the unloved... but he shall acknowledge the firstborn... by giving him a double portion of all that he has" (Deuteronomy 21:15-17, ESV). Such provisions promote familial stability, essential for capitalist productivity, as secure homes enable labor and innovation.

These Torah principles influenced American democracy by inspiring land reforms. Jefferson, drawing on biblical agrarianism, advocated for smallholdings to cultivate republican virtue, echoing Torah's anti-monopolistic ethos (Novak, 2002). In capitalism, this stewardship tempers profit with ethics, as seen in Jewish legal traditions prohibiting boundary removal: "You shall not move your neighbor's landmark" (Deuteronomy 19:14, ESV). This moral boundary protects property rights while fostering communal harmony, vital for democratic pluralism amid market competition (Forrest & Hirayama, 2018).

Christianity builds on Torah foundations, incorporating New Testament teachings that spiritualize property, emphasizing charity and eternal inheritance while affirming earthly ownership's role in societal order. This dual focus shaped America's capitalist democracy by integrating moral imperatives into economic freedom. Proverbs, shared with Judaism, highlights inheritance: "House and wealth are inherited from fathers, but a prudent wife is from the Lord" (Proverbs 19:14, ESV). This underscores homes as generational blessings, promoting stability that undergirds democratic participation—owners, rooted in communities, engage civically. The New Testament reframes property through Christ's teachings, viewing it as a tool for stewardship rather than hoarding. In the Parable of the Talents (Matthew 25:14-30), servants are entrusted with property to multiply, rewarded for faithful management: "Well done, good and faithful servant. You have been faithful over a little; I will set you over much" (Matthew 25:21, ESV). This endorses capitalist enterprise—investment and growth—while demanding ethical use, aligning with American ideals of opportunity tempered by responsibility. Acts 4:32-35 depicts early Christians sharing property voluntarily: "No one said that any of the things that belonged to him was his own, but they had everything in common" (Acts 4:32, ESV). Yet, this communality presupposes private ownership, as seen when Ananias sells land but lies about proceeds (Acts 5:1-4), affirming individual rights: "While it remained unsold, did it not remain your own?" (Acts 5:4, ESV).

4. Addressing the home ownership problems: a cross-ideological imperative for sustaining democracy and capitalism

Forrest and Hirayama (2018) observe similar patterns in home ownership across highly financialized economies – the United States, UK and Japan. They start off by pointing to the growth of individual home ownership in the post-war mature capitalist economies, which occurred within a society characterized by “the atomized, nuclear family; suburbanization; high growth; the conventional mortgage market and a young, working population”. They then argue that these conditions have changed and now they contrast between ‘real estate families’ or ‘accumulating families’ which maintain or further accumulate valuable multiple property assets over generations; ‘dissipating families’ which are forced to deploy and diminish their property assets accumulated in the exceptional era; and propertyless ‘perpetual renter families’ (Forrest and Hirayama, 2018). According to them, this leads to a social re-stratification driven by ownership of residential property and changing demographics which has a profound impact on social stability. The ensuing inequality with top quintiles owning about 80% of housing

wealth while bottom quintiles renting precariously (Burchard, 2025) may result in undermining the "property-owning democracy", where diffuse ownership stabilizes society by fostering stakeholding; instead, concentrated housing wealth may start eroding social cohesion (Vitali, 2023). Franc and Pavlović (2018) link such inequality to radicalization, finding that perceived economic disparities correlate with extremist attitudes. In the US context, this polarization has led to millennials and gen z feelings of a rigged system spilling over into envy and resentment of successful people (Elson and King, 2025). In a poll conducted by Elson and King (2025), millennials and gen z exhibit higher than average envy, relatively disagreeing with showing sympathy to very successful people failing, relatively taking joy in seeing very successful people failing and relatively disagreeing with the statement that people at the top usually deserve their high position. The processes of re-stratification and polarization may perpetuate a deeply rooted environment of resentment and envy which undermines a merit-based capitalist system functioning under a democratic political order. Koopmans (2014) extends this to religious fundamentalism, where economic exclusion heightens out-group hostility, potentially radicalizing disenfranchised youth into terrorist actions. Koopmans (2014) is possibly the only study to have tested and confirmed a multivariable relationship between home ownership and religious fundamentalism as well as out-group hostility, establishing that people who live in rented housing are significantly more likely to express and be influence by these.

As millennials perceive the system to be rigged by older generations against them and start pointing fingers, they are at risk of becoming disenfranchised which in turn threatens participatory governance. Some academicians such as Chaput (2020) documents and labels the process as "millennial democratic fatigue", pointing to a lagging voter turnout among millennials. Chaput (2020) argues that although millennial voter participation increases with time, by 2016 they had still been unable to break 50%, something that the boomers reached in their third election cycle and generation X achieved in the sixth election cycle. Millenials also show a rapidly declining trust in government and both the major political parties, which also translates in almost half of this generation's active voters registering as independents, significantly higher than the 39% of generation X and 37% of boomers (Chaput, 2020). Similarly, declining homeownership undermines capitalism's legitimacy, shifting views toward socialism. Faria (2022) argues Millennials' 51% favorable view of socialism stems from inequality perceptions, viewing capitalism as "rigged".

Addressing the home ownership problems can be a cross ideological imperative that builds consensus across the political spectrum as it spans over a wide nexus of issues and solutions that can be prioritized accordingly – income and wealth inequality, property rights and increasing supply through deregulation and/or state support, tax incentives and public-private partnerships, boosting youth engagement and participation, even promoting environmental sustainability. Beyond the binary of Democrats and Republicans, U.S. politics comprises diverse voter coalitions: progressives and democratic socialists on the left; centrists and moderates bridging the center; fiscal and social conservatives on the right; libertarians advocating minimal government; populists on both flanks emphasizing anti-elitism; and environmentalists or Greens prioritizing sustainability. Each group has stakes in resolving homeownership barriers, as failure to do so risks eroding their core constituencies and ideological legitimacy.

For progressives and democratic socialists—core to the Democratic Party's left wing, including supporters of figures like Bernie Sanders and Alexandria Ocasio-Cortez—addressing homeownership is vital to combating what is perceived as systemic inequality. These groups view housing as a human right and oppose it being transformed into a mere commodity or asset investment type. They see declining ownership as a manifestation of

neoliberal failures that widen racial, gender and class divides (Isom, 2024). Importance lies in their commitment to social justice: progressive Millennials and Gen Z face barriers like student debt and wage stagnation, perpetuating cycles of poverty and radicalizing youth toward extreme anti-capitalist views (Franc & Pavlović, 2018; Faria, 2022). While this group often praises aspects of socialism and even communism, the political leaders realize that they cannot let a substantial demographic group slip into genuinely opposing the capitalist future of the USA as this will open a path for conservatives to further brand liberals as too radical and tighten their grip on the moderates. By prioritizing policies that effectively assist young generation in attaining home ownership, progressives can retain young voters, who prioritize wealth-building through homeownership. Such policies may include progressive taxation on high-value properties, zoning deregulation and limiting the ability of corporations to acquire large number of homes as investment assets that could be popular across the isles with very few exceptions (for instance, radical libertarians).

Centrists and moderates, spanning establishment Democrats and Republicans (e.g., Biden-Harris supporters and Never-Trump conservatives), emphasize pragmatic, evidence-based policies to foster economic mobility. For them, homeownership's decline threatens the middle-class consensus that sustains bipartisan governance, as it erodes consumer spending and local tax bases. Importance stems from their focus on stability: Gen Z's delayed ownership correlates with lower workforce participation and innovation, hindering growth. Feasibility lies in cross-party reforms. Bipartisan efforts, like the Senate's housing supply bills, propose tax credits for affordable construction and streamlined permitting.

Fiscal and social conservatives, underpinning the Republican Party (e.g., evangelical voters and small-government advocates), see homeownership as central to family stability and self-reliance. The crisis is important because it undermines traditional values: Delayed marriage and children among Gen Z correlate with lower ownership, threatening demographic renewal and moral fabric. Conservatives view this as a failure of overregulation, alienating working-class voters who feel squeezed by inflation. Solutions fit conservative principles through deregulation and incentives; reducing zoning restrictions to boost supply is a key proposal; tax reforms, like expanding mortgage interest deductions for families, promote ownership without subsidies. Social conservatives could support community-based initiatives, such as faith-group housing co-ops, aligning with pro-family policies, which in turn may found support from both moderates and progressives.

Libertarians, often aligning themselves with Republican party fringes, would favor minimizing government intervention in the housing sector in order to boost housing supply and thus push prices down. The current housing affordability crisis is viewed by them as a product of government overreach, which could be fueled by good intentions, but has ultimately stifled the ability of entrepreneurs to provide sufficient supply for the increasing demand. Furthermore, libertarians would be concerned to grow young voter cohorts being radicalized against property rights and capitalism as they cannot root their careers and personal development in home ownership. The libertarian camp is likely to prefer the abolition of zoning laws in order to allow vertical development and market-driven supply, as well as radical state-level deregulation. Libertarians may reach across the isle finding support for increasing land and estate taxes in order to shift the tax burden away from income as they often deem personal income taxes to be unconstitutional, highly ineffective and unjust.

Populists in the camps of both Sanders and Trump would view the home ownership inequality and home affordability crisis as elite-driven unjust policies that disenfranchise youth and less affluent voters. Housing can be politically exploited by populists who thrive on division of “us” (workers, youth, less affluent people) versus “them” (affluent elites, multinational investors in American real estate, corporations thriving on cheap money

inaccessible for the masses). Addressing this can be a unifying issue that provides high political benefits to populists on both sides of the isle who seek to tone down political rhetoric and steer away from radicalization that has recently spurred bursts of political violence - for instance, Cenk Uygur's relationship with Charlie Kirk and Bill Maher's dinner with President Trump. While these camps may disagree on solutions such as curbing immigration and rent control, they may coalesce on a cross-ideological push for limiting corporations in investing in residential real estate, higher taxes on high value estate, zoning reform and construction deregulation in order to increase supply.

5. Conclusion

Cross-ideological political solutions are essential to preserving youth trust in the American system of democracy and liberal capitalism. Disenfranchised and radicalized cohorts of generation z and millennial voters who suffer from the housing affordability crisis are a threat to all major political camps in the US two party system. At the same time, the problem itself and many possible solutions touch on topics that are wildly popular across the political spectrum – equity, markets, family, liberty, anti-elitism. In times of heated rhetoric, perpetual partisan bickering and aggressive deployment of institutions against the political opponent, cool-headed politicians and pundits seek to reach across the isle in order to steer the US further from the outbursts of political violence and back to an effective debate and governance. Boosting home ownership for the younger generations may be a prime opportunity of enormous political benefits and very little costs.

As a final note, the US home affordability problems are not the direst among the developed economies with changes in real home prices and disposable income diverging much more drastically over the last half a century in Australia, the UK, Ireland, and Canada, for example.

References

1. Aristotle. (2009). Politics (B. Jowett, Trans.). Project Gutenberg. https://www.gutenberg.org/files/6762/6762-h/6762-h.htm (Original work published ca. 350 B.C.E.)
2. Arouet, M. [@MichaelAArouet] (2025) Real home prices and disposable income. The US is in fairly good shape. Things in some other countries are simply wild. [Image attached] [Post] X. https://x.com/MichaelAArouet/status/1953122660088418587
3. Autor, D. H. (2014). Skills, education, and the rise of earnings inequality among the "other 99 percent". Science, 344(6186), 843-851.
4. Bass, D. (2025) Poll: boomers, millennials blame each other for housing affordability crisis. https://www.carolinajournal.com/poll-boomers-millennials-blame-each-other-for-housing-affordability-crisis/
5. Chaput, W. M. (2020). Democratic Apathy: Exploring the Roots of Millennial Democratic Fatigue.
6. Coates, B., Moloney, J. and Clay, K. (2022) Do millennials really have it tougher in the housing market? https://grattan.edu.au/news/do-millennials-really-have-it-tougher-in-the-housing-market/
7. de Tocqueville, A. (1835). Democracy in America (H. Reeve, Trans.). Project Gutenberg. https://www.gutenberg.org/files/815/815-h/815-h.htm
8. Economic Policy Institute (2025). The productivity-pay gap. https://www.epi.org/productivity-pay-gap/
9. Elson, C., & King, K. (2022). Views of democracy, egalitarianism and capitalism among millennials and Gen Zs. Wake Forest University Center for the Study of Capitalism.
10. Faria, A. (2022) Comrade snowflake? The Independent Review, v. 27, n. 2, Fall 2022, ISSN 1086–1653
11. Faria, A. B. (2022). Comrade snowflake? Why millennials won't be socialists forever. The Independent Review, 27(2), 243–271.

12. Forrest, R., & Hirayama, Y. (2018). Late home ownership and social re-stratification. Economy and Society, 47(2), 257–279. https://doi.org/10.1080/03085147.2018.1459368
13. Forrest, R., & Hirayama, Y. (2018). Late home ownership and social re-stratification. Economy and Society, 47(2), 257-279.
14. Franc, R., & Pavlović, T. (2018). Systematic review of quantitative studies on inequality and radicalisation. Dialogue About Radicalisation and Equality.
15. Gregory, V. (2023). Generational Gaps in Income and Homeownership. Economic Synopses, (15).
16. Gregory, V., & Bloodworth, K. (2023). Understanding the Generational Gaps in Homeownership. Economic Synopses, 24.
17. Hamptons (2024) Millennials' mortgage misfortune. https://www.hamptons.co.uk/articles/millennials-face-rising-mortgage-costs-in-second-half-of-the-loan#/
18. Isom, K. A. (2024). Home Ownership Inaccessibility for Upcoming Generations in the United States. Ballard Brief, 2024(3), 3.
19. Khadem, N. (2023) Boomers, Generation X or Millennials: Who has it worse when it comes to buying a home and paying it off? https://www.abc.net.au/news/2023-02-06/baby-boomers-generation-x-millennials-housing-interest-rate-rise/101929468
20. Koopmans, R. (2015). Religious fundamentalism and hostility against out-groups: A comparison of Muslims and Christians in Western Europe. Journal of Ethnic and Migration Studies, 41(1), 33-57.
21. Locke, J. (1689). Second treatise of government. Project Gutenberg. https://www.gutenberg.org/files/7370/7370-h/7370-h.htm
22. Macdonald, D. (2015) The wealth advantage. https://www.policyalternatives.ca/wp-content/uploads/attachments/Wealth_Advantage.pdf
23. Madison, J. (1787). Federalist No. 10. In A. Hamilton, J. Madison, & J. Jay, The federalist papers.
24. Madison, J. (1792). Property. In The papers of James Madison. University of Chicago Press.
25. Moloney, J. and Coates, B. (2022) The housing game has changed – Millennials have it harder. https://grattan.edu.au/news/the-housing-game-has-changed-millennials-have-it-harder/
26. NAR (2024) Highlights from the profile of home buyers and sellers. https://www.nar.realtor/research-and-statistics/research-reports/highlights-from-the-profile-of-home-buyers-and-sellers
27. Phillips, J. (2018). American agrarianism's answers to the nation's (in)securities. Connecticut Public Interest Law Journal, 9(2).
28. Pisano, N. (2025) 2025 data: surprising findings on homeownership by generation. https://cleveroffers.com/research/home-ownership-by-generation/?utm_source=press+release&utm_medium=pr&utm_campaign=generational_views_2025
29. Sarna, A. (2025) Housing peaks & market concentration. https://offthecharts.substack.com/p/housing-peaks-and-market-concentration
30. The national association of home builders (2025) Nearly 60% of US households unable to afford a $300k home. https://www.nahb.org/blog/2025/03/priced-out-affordability-pyramid
31. Vitali, J. (2023). The Property Owning Democracy. Policy Exchange.

Artificial Intelligence in Managing Forced Migration: Evidence from Europe

Vasil Pavlov

University of National and World Economy, Sofia, Bulgaria
vpavlov@unwe.bg

Abstract. Migration has emerged as a key challenge for European Union (EU) member states, particularly in light of the growing number of national and regional military conflicts, political persecutions, and persistent security shortcomings. This paper investigates the potential implications of artificial intelligence (AI) in managing forced migration within the EU. Employing a quantitative, data-driven research, the study analyzes basic characteristics of forced migration to the EU over the past decade and identifies best practices from MS that have effectively integrated AI-based tools into their migration management systems. In this regard, demographic data on asylum seekers arriving in the EU between 2015 and 2024 are extracted from Eurostat, while literature review and official public institutional data are utilized to examine the current applications of AI in migration management. The central research question explores whether AI can serve as a reliable and effective instrument for managing increasingly complex and mixed migration flows.

Keywords. Artificial Intelligence, Forced Migration, Asylum Seekers, Europe.

1. Introduction

Lately, the world has faced a growing number of conflicts, leading to significant global instability [1]. Regardless of their intensity, these instabilities frequently force people to flee their countries of origin in a search of a safer and better living conditions. The EU, due to its relatively supportive social and economic conditions and its established legal framework and asylum seekers' system, has become a primary destination for many migrants over the past decade.

A huge increase of asylum applications in the EU has been observed in the period 2014-2016 [2], which was mainly motivated by the civil war in Syria, part of the broader early 2010s Arab Spring protests, that emerged from widespread discontent with the Syrian government. Initially peaceful demonstrations against the President Bashar al-Assad escalated into violent repressions, and brutal armed rebellion, that lasted almost 14 years with the fall of the Assad regime [3]. This period marked what became widely referred to as the European migration crisis.

However, after a relative decline in asylum inflows between 2017 and 2021, immigration to the EU have again increased substantially, primarily triggered by the Russian invasion of Ukraine in February 2022 [4], as well as by other geopolitical and regional instabilities, such as the Taliban's takeover in Afghanistan [5] and periodically escalating tensions between Israel and Palestine, especially in the aftermath of the October 2023 Hamas attack [6]. These overlapping crises continue to challenge the EU's capacity to respond relevantly to ongoing humanitarian needs and migration pressures.

This paper has two primary objectives. First, it aims to examine the forced immigration trends to the EU over the past decade (2015–2024) through the extraction and analysis of

official data from Eurostat and Statista. Second, it seeks to evaluate the implications of current AI instruments in managing complex mixed migration flows. This evaluation will be conducted by gathering information from official sources such as the European Border and Coast Guard Agency (Frontex) and the European Parliament, as well as through a comprehensive review of relevant academic literature. The main research question put in this study is whether AI-based approaches can serve as reliable tool for managing forced migration inflows. By examining the potential applications of AI-based decisions in complex issues such as migration, this study aims to contribute to a broader understanding of the advantages of adopting such technologies, particularly in less developed EU member states.

2. Literature Review

Over the past few years, a rapid development and a growing use AI technology has significantly transformed many sectors, including healthcare, finance, education, and public governance, just to name a few. As AI instruments become widely integrated into decision-making processes, their application has expanded beyond coping commercial needs to areas, such as migration management.

This widespread use of AI has attracted the attention of scholars worldwide, as many recent studies examine the integration of AI-based tools into forced migration management, raising questions about both its potentials and its risks. The escalating interest towards this issue is presented in Rönsch's research, which offers a systematic literature review of AI applications in EU migration governance, highlighting the benefits of efficiency, however, expressing warnings about unregulated deployments that may compromise migrant rights [7]. On the other hand, Nalbandian proves that sophisticated AI tools have been used to register and manage vulnerable populations such as migrants and asylum seekers, however, without paying attention to "potential misuses of the data collected and the overall ethical and legal underpinnings of these operations" [8].

Legal and ethical concerns are expressed by Yang, Zuiderveen Borgesius, Beckers, and Brouwer who identify major risks associated with AI at EU borders, including threats to privacy, fairness, and legal recourse [9]. On the other hand, another study investigates the advantages of: firstly, predictive AI, aimed at the governments support in time estimation, destination and size of refugee inflows, and, secondly, the Natural Language Processing (NLP) integration within a broader repertoire of techniques to automate immigration systems. However, the paper suggests that to better harness the analytical power of AI, these tools should be recognized as "inherently political, in the sense that they enshrine a specific view of power and relations of subordination" [10]. Regulatory critiques, such as those emerging around the EU AI Act, argue that exemptions for migration-related applications undermine accountability [11].

Case-specific studies also raise critical points. For example, Molnar discusses how specialized accountability frameworks are essential to protect migrants' fundamental rights, including freedom from discrimination, privacy protections, and procedural justice guarantees such as access to impartial decision-makers and the right to appeal [12].

Across this literature, a consensus emerges: AI tools may support migration governance, but only when grounded in robust legal, ethical, and participatory frameworks.

3. Current Asylum Situation Analysis

As previously highlighted in the very introduction of the current paper, the contemporary world is confronted with a significant rate of crises and security challenges. The pushing factors that make people flee their countries of origin have different nature - they vary

from climate change, political instability, social inequalities, to poor economic conditions. In addition to these underlying causes, armed conflicts and other forms of violent unrest, persecution, etc. remain significant catalysts for forced displacement. To contextualize this phenomenon, Figure 1 presents data on ongoing conflicts worldwide, categorized according to their varying levels of intensity.

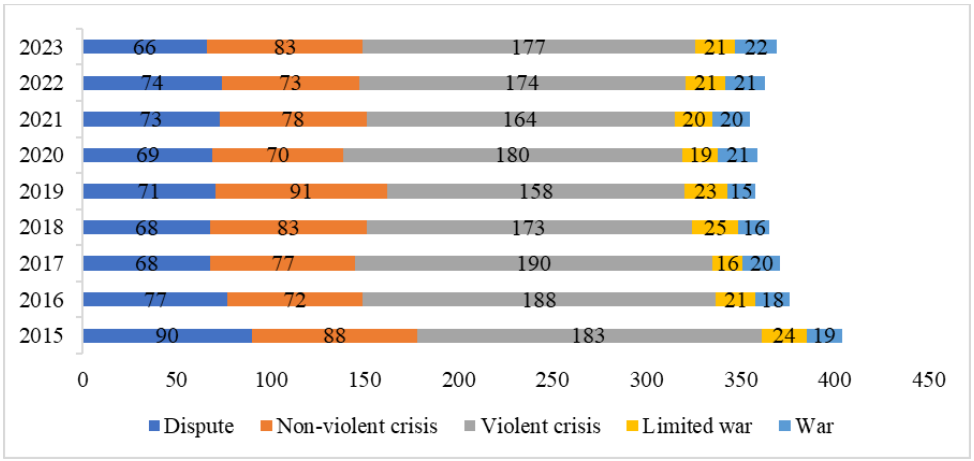


Figure 1. Number of conflicts worldwide from 2015 to 2023, by conflict intensity (Source: Statista [1])

As illustrated in Figure 1, the period under study reveals that more than 350 conflicts of varying intensity occur all over the world on an annual basis. These conflicts, at the roots of which stand variety of political, cultural, ethnic, territorial, and ideological contradictions lead to forced displacement of people. Europe, due to its better living standards and regulated immigration policies, is attractive destination.

In addition to the data in Figure 1, it should be noted that two peaks in conflict occurrence can be observed: the first in 2015, connected with the onset of the so-called European migration crisis for which it was already mentioned, and the second during the 2022–2023 period, marked by renewed and/ or newly triggered conflicts, such as the Russian invasion of Ukraine and some tensions in the Middle East and North Africa (MENA) region like between Israel and Palestine. These surges in the number of conflicts worldwide, as well as their intensity, correlate with the inflows of asylum seekers in the EU. The latest is further supported and confirmed by the data in Figure 2, which represents the number of first-time asylum applications submitted at the EU level on an annual basis between 2015 and 2024. Moreover, the comparison between two indicators demonstrates a direct relationship between rising global instability and the demand for international protection within the EU.

As visible from the data in Figures 1 and Figure 2 there is a relation between the number of ongoing conflicts between 2015 and 2024 and the volume of asylum applications, submitted within the EU in the same period. In other words, increased global violence tend to reflect in higher numbers of people seeking international protection in Europe. This trend, however, put the EU in a position, which requires the decision-makers and responsible public institutions to balance between fundamental human rights protection, on the one hand, and to keep the EU borders safe, preventing irregular immigration, on the other. Ensuring border security is a key priority, as some immigrants may pose security risks, including potential radicalization, terrorism, or extremism, however, it is not the only element of migration management system.

Once they have arrived and settled in the territory of the member states, the institutions in charge face other, no less important challenges such as social, economic and cultural integration. In this regard, some of the main advantages and disadvantages of applying AI-based methods in the management of mixed migration flows are presented in the next section.

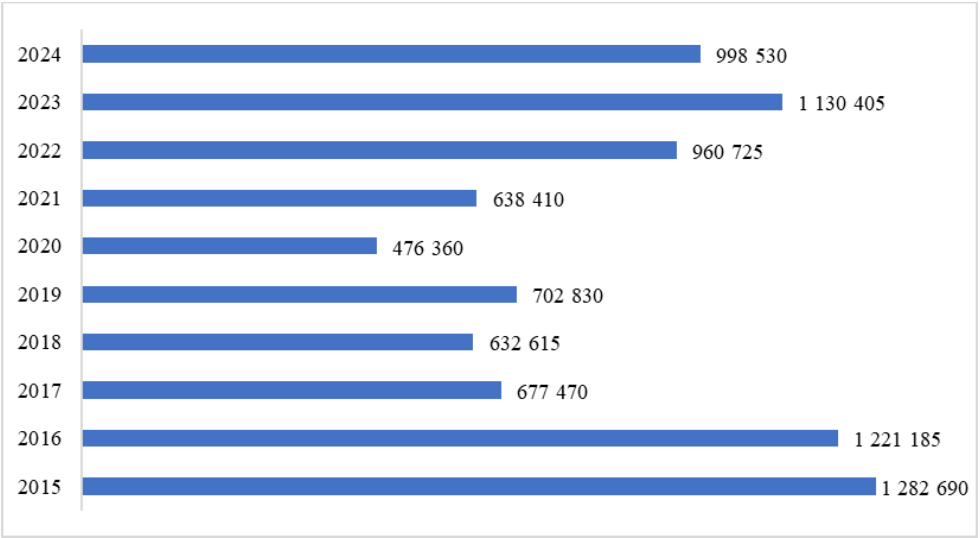


Figure 2. First-time asylum applications in the EU – 27 countries (from 2020), presented annually for the period 2015-2024 (Source: Eurostat [2])

4. Benefits and Limitations of AI Technologies in Managing Forced Migration

In the contemporary technologically advanced world, further sophisticated due to the AI era, the EU has increasingly turned to such solutions to enhance border management, as also clarified by the literature review. In addition, according to an analysis by the European Parliamentary Research Service, the EU employs AI-driven systems in fields such as: biometric identification; emotion detection; algorithmic risk assessment; and migration monitoring, data analysis, and forecasting [13]. In addition to border control protection, Frontex adds that they use drones that are equipped with advanced AI technologies such as sensors and communication systems, offering real-time surveillance that significantly enhances situational awareness [14]. These innovations reflect the EU's broader strategy to implement data-driven and intelligent decision-making processes in managing migration and ensuring border safety. However, some decisions face critique among academic society. For instance, Kontak argues that Frontex fails in its legal obligation to respect fundamental rights in relation to its biometric policy [15]

Despite growing concerns about ethical and legal risks, several good practices have been identified across the literature regarding the use of AI in migration management. First, predictive AI tools have shown potential in supporting governments by providing early estimations of refugee flows, including their timing, destination, and scale. Such tools can improve preparedness and allocation of resources, enabling more responsive and proactive migration governance [10].

Another positive application involves the integration of NLP to assist in automating administrative procedures in immigration systems. NLP can facilitate communication with asylum seekers and streamline data collection and case processing, especially in multilingual or high-volume contexts [10].

Further good practices include the use of AI-powered registration and population management systems that help document and track vulnerable groups, such as as ylum seekers and displaced individuals. When properly designed, these systems can enhance operational efficiency, improve access to services, and support identification processes in humanitarian settings [8].

Importantly, several authors stress that these technologies must be implemented within transparent and rights-based accountability frameworks. Molnar highlights the need of implicating procedures to protect fundamental rights, such as privacy, freedom from discrimination, and procedural justice, particularly the right to a fair decision-maker and access to appeals mechanisms [12].

Lastly, the literature review outlines that the application of AI-based decisions in this specific field should be issued with an awareness of their inherently political nature, involving the recognition of how algorithmic systems may reflect and reinforce existing power dynamics. Thus, good practice requires incorporating participatory approaches in the design and deployment of AI tools, ensuring input from affected communities, legal experts, and civil society actors [10].

Last but not least, it should be stressed that the reviewed literature does not encompass all possible and existing applications of AI in the field of migration management; rather, it highlights selected cases and trends that have been the focus of recent academic inquiry.

5. Conclusion

The turbulent security environment, marked by many conflicts with varying intensity over the last decade, force millions of people to flee their country of origin. More often than not, forcibly displaced individuals move to safer countries, which offers them better social and economic conditions. The current analysis confirms that surges in violent conflicts directly correlate with incensements in the volume of asylum applications in the EU. In this regard, AI-based solutions have emerged as a reliable instrument in migration management systems.

By critically reviewing official institutional sources and academic literature review reviles that AI technologies, such as biometric identification, predictive analytics, and automated surveillance, just to name a few, offer substantial potential in migration management process. When properly designed and regulated, these tools can strengthen institutional capacity to forecast migration flows, optimize administrative procedures, and facilitate the protection of vulnerable populations.

This paper also highlights critical concerns related to ethics, legality, and accountability. It also becomes clear that AI application that is not enough regulated has the potential to harm some fundamental rights of migrants. Therefore, clearly defined mechanisms and well-developed legal frameworks are urgently needed for effective AI deployment.

In conclusion, while AI represents a reliable alternative for EU migration management modernization, its benefits can only be fully applied through a balanced approach. This should be done in accordance with technological innovation that respect human rights and democratic values. Last but not least, a successful integration of AI-based decisions in migration management will also generate new opportunities for business innovation, particularly in developing data-driven solutions, security technologies, and digital platforms that support sustainable and ethical migration governance.

Acknowledgements

This work was financially supported by the UNWE Research Programme under the project "Application of Artificial Intelligence for Sustainable Business Development in the Context of Climate Change" (Research Grant No. NID NI -35/2025/A).

References

1. Statista, Number of conflicts worldwide from 2005 to 2023, by conflict intensity, 2025, (https://www.statista.com/statistics/262935/conflicts-worldwide-by-intensity/)
2. Eurostat, Asylum applicants by type, citizenship, age and sex - annual aggregated data, 2025, (https://ec.europa.eu/eurostat/databrowser/view/migr_asyappctza/default/table?lang=en)
3. Al Jazeera, The end of Syria's forever war, 2024, (https://www.aljazeera.com/gallery/2024/12/10/the-end-of-syrias-forever-war)
4. Operational Data Portal, Ukraine Refugee Situation, 2025, (https://data.unhcr.org/en/situations/ukraine)
5. Council on Foreign Relations, The Taliban in Afghanistan, 2025, (https://www.cfr.org/backgrounder/taliban-afghanistan)
6. Al Jazeera, Updates: Israel's genocide continues in Gaza two years since start of war, 2025, (https://www.aljazeera.com/news/liveblog/2025/10/7/live-israels-genocide-continues-across-gaza-two-years-since-start-of-war)
7. Rönsch, Anna-Marie, The Use of Artificial Intelligence in Migration Management -A Systematic Literature Review, 2024, (https://ssrn.com/abstract=4953245)
8. Nalbandian, L., An eye for an 'I:' a critical assessment of artificial intelligence tools in migration and asylum management, 2022, (https://doi.org/10.1186/s40878-022-00305-0)
9. Yang, Y. et al., Automated decision-making and artificial intelligence at European borders and their risks for human rights, 2024, (https://doi.org/10.48550/arXiv.2410.17278)
10. Iazzolino, G., Trading Efficiency for Control: the AI Conundrum in Migration Management, 2025, (https://doi.org/10.5130/ccs.v17.i1.9423)
11. Stewart, L. S., The regulation of AI-based migration technologies under the EU AI Act:(Still) operating in the shadows?, European Law Journal, 30(1-2), 122-135, (https://doi.org/10.1111/eulj.12516)
12. Molnar, P., Technology on the margins: AI and global migration management from a human rights perspective, Cambridge International Law Journal, 8(2), 305-330, (https://doi.org/10.4337/cilj.2019.02.07)
13. European Parliamentary Research Service, Artificial intelligence at EU borders Overview of applications and key issues, 2021, (https://www.europarl.europa.eu/RegData/etudes/IDAN/2021/690706/EPRS_IDA(2021)690706_EN.pdf)
14. Frontex, Last Month in the Field – May, 2025, (https://www.frontex.europa.eu/media-centre/news/news-release/last-month-in-the-field-may-Cxgu6T#:~:text=The%20drones%20are%20equipped%20with,and%20discuss%20its%20future%20potential)
15. Kontak, M., Biometric Borders Envisaged by Frontex: Fundamental Rights in the Backseat, European Papers-A Journal on Law and Integration, 2024(2), 621-640, (https://doi.org/10.15166/2499-8249/773)

Regional Focus and Sentiment Trends in Bulgarian Media: A Study Based on Web Publications

Plamen Milev

University of National and World Economy, Sofia, Bulgaria
pmilev@unwe.bg

Abstract. The digital media landscape in Bulgaria is a complex ecosystem where regional issues often struggle for visibility against national political narratives. This study investigates regional focus and sentiment trends within Bulgarian web publications using a media intelligence approach. Based on a large-scale dataset comprising monthly word frequency records gathered from 8 major web platforms over a one-year period, the research applies advanced Natural Language Processing (NLP) techniques. Data collection was performed via automated web scraping, and analysis was conducted using a custom Python script. We analyze the distribution of media attention across Bulgaria's 28 administrative districts. The findings reveal a significant "Sofia-centric" bias and highlight how regional sentiment is shaped by local socio-economic challenges and global event-driven fluctuations.

Keywords. Media Intelligence, Text Analytics, Web Scraping, NLP, Web Publications.

1. Introduction

Since the beginning of the transition in 1989, the Bulgarian media landscape has evolved from state instrumentalization to a state of media capture by political and oligarchic interests [1]. This transition has not been linear, but it has been marked by a shift from direct government control to a more complex, indirect system of influence where editorial independence is often secondary to the strategic interests of media owners. In this environment, the relationship between media and politics remains a central factor in shaping public discourse [2]. The concentration of media ownership and the lack of transparency in funding have significantly narrowed the range of viewpoints available to the public, leading to a polarized media environment where national narratives are frequently synchronized with political agendas. While digital platforms have expanded the reach of information, offering a seemingly endless stream of real-time updates, regional media outlets continue to struggle with professional and technological contradictions, particularly in the quality of visual journalism [3]. Rapid digitalization has created a noticeable digital divide. While major national outlets in Sofia invest in high-end production and sophisticated digital storytelling, local newsrooms often lack the resources, training, and infrastructure to compete. This technological lag is often accompanied by a decline in professional standards, as regional journalists face increased pressure to produce high-volume content with minimal staff, often resulting in a reliance on press releases or sensationalist framing to maintain audience engagement.

This paper presents a data-driven analysis of the Bulgarian media landscape, focusing on regional representation and sentiment. The necessity for such an analysis comes from the growing concern over the uneven distribution of information across the country. By quantifying the geographic focus and identifying "information deserts" – regions that are chronically underrepresented or ignored by the mainstream media – we aim to understand how localized narratives vary across administrative districts and how they reflect the broader socio-

economic and political health of the country. Furthermore, this study seeks to move beyond qualitative observations by employing media intelligence and NLP techniques. In a captured media environment, traditional content analysis is often insufficient to capture the subtle patterns of bias and neglect. By analyzing 248,122 web publications across eight major platforms, we can objectively measure the "Sofia-centric" bias and determine whether the emotional tone of news varies based on the economic prosperity or political significance of a given region. This research provides an empirical foundation for discussing the role of media in regional development and the health of local democracy in Bulgaria.

2. Related Work

The analysis of Bulgarian media requires a multi-layered review of the political context, thematic narratives, and technological capabilities for text processing. The structural environment of Bulgarian media is characterized by high degrees of ownership concentration and limited plurality [4]. This environment often leads to inadequate coverage of complex topics, a challenge shared by other Balkan nations where media distrust is frequently fueled by a lack of effective communication regarding institutional functions [5]. Specific narratives often dominate the digital space, frequently utilizing negative sentiment to drive engagement. A specialized framework for analyzing disinformation narratives highlights the vulnerability of the Bulgarian context to misleading information regarding sensitive social groups [6]. National reports on online hate speech further demonstrate the prevalence of toxic content in digital forums [7], while the representation of migration along the Balkan route shows significant verbal and visual biases [8]. These dynamics influence the public's perception of civil society organizations, which are often portrayed through conflicting positive and negative portraits [9]. The technical feasibility of large-scale media monitoring relies on progress in Bulgarian NLP. Foundational overviews of text analytics resources for the Bulgarian language provide the necessary framework for data processing [10]. Specialized multi-class classifiers have been developed to handle complex tasks such as detecting toxicity in news articles [11], which is crucial for identifying inflammatory regional reporting. The production of news has adapted to digital pressures, with a global trend toward longer and more negative headlines designed to capture audience attention [12]. Supporting this digital ecosystem are integrated technological systems, such as those combining Enterprise Content Management (ECM) and Business Process Management (BPM) [13], alongside the systematic classification of process data types [14]. Furthermore, the technological features of modern video-sharing platforms [15] and the optimization of software testing via AI [16] provide the infrastructure necessary for sophisticated media analysis. Regional media play an essential role in tracking sustainable development. Text mining techniques have been successfully applied to identify sustainability trends in regional publications [17]. Comparative analyses of "digital voices of sustainability" [18] underscore the importance of automated sentiment analysis in understanding how public perceptions of green policies vary across borders.

While the existing literature provides a robust theoretical foundation for understanding media capture and the technical challenges of Bulgarian NLP, there is a clear need for large-scale empirical studies that map these dynamics geographically. This research bridges that gap by moving from conceptual frameworks to a high-volume, longitudinal analysis. To achieve this, a specialized technical infrastructure was developed, utilizing automated web scraping and custom analytical tools. The following section details this methodology, describing the data harvesting process from eight major platforms and the Python-based workflow used to transform raw word frequencies into regional and sentiment-based insights.

3. Methodology

The study employs a media intelligence pipeline applied to a dataset of 248,122 web publications collected from 8 major Bulgarian web platforms: Bulgaria ON AIR, Bloomberg, BNR, BNT, bTV, DW, Eurocom, and Nova. These platforms were selected to ensure a representative cross-section of the Bulgarian media landscape, including public broadcasters, private national television networks, and international news agencies operating within the local digital market. The methodological approach proposed in this research includes the following phases (Fig. 1):

- Data collection via web scraping: the data was harvested using automated web scraping techniques. The scraping engine, developed specifically for this longitudinal study, targeted the selected news platforms by monitoring their web publications. This mechanism allowed for the systematic extraction of textual content while minimizing technical overhead. The collection covered a continuous 12-month period, from April 2024 to March 2025, ensuring that the findings reflect both cyclical seasonal patterns and specific event-driven fluctuations in the national agenda.

- Linguistic processing and NER proxy: the primary analysis was conducted using a custom Python script designed to process the large-scale CSV word frequency datasets. This script performed several preprocessing tasks, including the removal of stop words and the normalization of Cyrillic text. Given the high volume of data, a Named Entity Recognition (NER) proxy was implemented within the script to isolate and count mentions of Bulgaria's 28 administrative districts and their main urban centers. By utilizing foundational text analytics resources for the Bulgarian language, the script was able to map the geographic distribution of news coverage, identifying which regions received the highest editorial priority.

- Sentiment proxy classification: sentiment is analyzed through thematic keyword clusters rather than traditional sentence-level analysis, allowing for high-efficiency processing of large-scale lexical data. Following established models for toxicity and sentiment detection in Bulgarian news, the script applied a lexical classifier. Negative (crisis) clusters include high-frequency terms like "криза" (crisis), "война" (war), and "проблем" (problem), which often signal political or social instability. Conversely, positive (development) clusters include terms such as "успех" (success), "туризъм" (tourism), and "инвестиция" (investment), typically associated with regional growth and economic stability.

- Temporal analysis: finally, the custom script performed a temporal analysis to track the monthly ranking and frequency spikes of high-profile political and geopolitical keywords. This allowed the research to identify event-driven fluctuations, such as sharp increases in mentions of specific political figures or international conflicts, and correlate these spikes with regional editorial shifts. This phase was crucial for determining how global news cycles interact with and potentially overshadow local regional narratives.

This technical infrastructure provides the necessary scalability and precision to process the extensive dataset of 248,122 web publications gathered across the eight selected platforms. By transforming raw, unstructured word frequency data into a structured set of regional and sentiment-based indicators, the media intelligence pipeline enables a granular examination of the Bulgarian digital ecosystem.

The following section presents the empirical results derived from this methodology, illustrating the clear disparities in regional media focus and the emotional valence associated with different news categories during the analyzed twelve-month period.

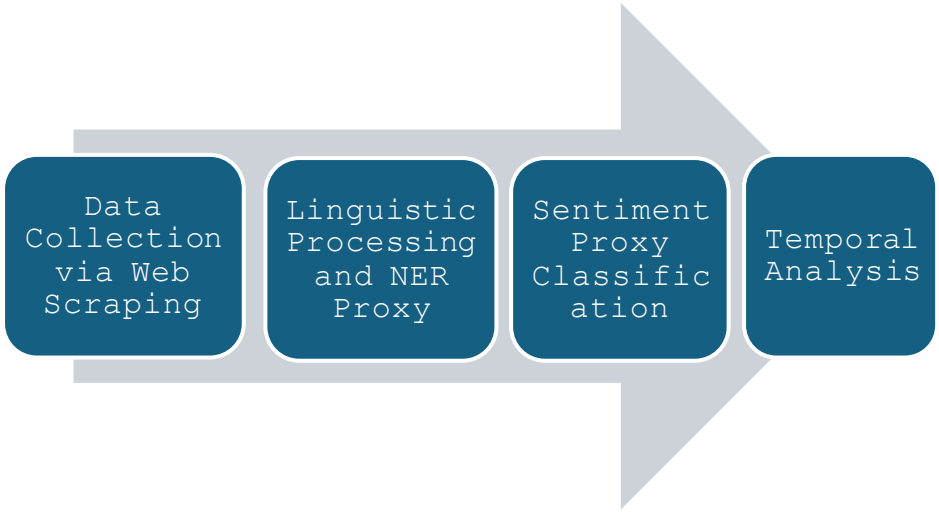


Fig. 1. Methodological approach

4. Results and Discussion

The analysis of the 12-month dataset reveals a persistent regional and thematic imbalance in the Bulgarian digital media ecosystem. This imbalance manifests in two primary ways: a geographic concentration of news coverage in the capital and an editorial agenda dominated by high-intensity geopolitical and national political crises at the expense of regional development topics. In terms of regional focus analysis, the distribution of media attention across major Bulgarian cities shows a massive concentration in the capital. Table 1 summarizes the results processed by the analysis script, revealing that Sofia-related news items account for nearly two-thirds of the total regional discourse.

Table 1. Regional focus in Bulgarian media (monthly average)

Region (city)	Average monthly mentions	Relative share (%)
Sofia (capital)	712.5	64.2%
Plovdiv	102.4	9.2%
Varna	94.8	8.5%
Burgas	81.2	7.3%
Ruse	38.6	3.5%
Others (23 districts)	80.5	7.3%

[Source: Author's data obtained from web scraping.]

This data confirms the existence of a profound geographic bias in Bulgarian digital media. While Plovdiv, Varna, and Burgas maintain a secondary tier of visibility due to their economic and administrative importance, the remaining 23 districts are virtually invisible in the national information flow, sharing only 7.3% of the total mentions. This quantitative evidence supports the "information desert" hypothesis, suggesting that peripheral regions are systematically marginalized in the digital public sphere.

In terms of sentiment and agenda volatility, the media agenda is heavily influenced by geopolitical events and political instability, as reflected in the high frequency of negative and crisis-related terms (Table 2).

Table 2 Frequency of sentiment-related keyword categories (monthly average)

Category	Representative keywords	Average frequency	Trend context
Crisis and negative	криза (crisis), война (war), проблем (problem), протест (protest)	485,2	High volatility during elections
Geopolitical	Украйна (Ukraine), Русия (Russia), Израел (Israel), Хамас (Hamas)	534,6	Sharp peaks during conflict events
Political transition	Тръмп (Trump), избори (elections), правителство (government)	682,1	Peaks in November 2024 and February 2025
Development	успех (success), туризъм (tourism), проект (project)	114,6	Seasonal peaks in summer months

[Source: Author's data obtained from web scraping.]

The keyword frequency data indicates that the Bulgarian digital audience is exposed to a high volume of crisis-driven content. The political transition category shows the highest average frequency, particularly driven by the peaks in November 2024 and February 2025, which correspond to the elections in the United States and domestic governmental shifts. In contrast, development keywords – which represent constructive regional narratives – have the lowest frequency, appearing roughly six times less often than political transition terms.

The dominance of Sofia reinforces the media capture narrative [1], where national political struggles often overshadow regional development needs. When regional news does appear, it is frequently framed through the lens of national crisis or administrative failure, rather than local potential. This concentration of attention suggests that the media acts more as a megaphone for central political power than as a platform for diverse regional voices. Furthermore, the peak in mentions of "Тръмп" (Trump) and other geopolitical actors in late 2024 and early 2025 demonstrates how international political cycles dominate the Bulgarian media space. This trend is often accompanied by negative emotional valence in headlines [12], as outlets compete for clicks using sensationalist and inflammatory language. The low visibility of development keywords highlights a missed opportunity for media to foster constructive regional discourse, pointing to a systemic preference for tabloid politics over substantive reporting on local sustainability and growth.

5. Conclusion

By integrating different research perspectives and analyzing a robust dataset from 8 prominent media platforms via web scraping and Python-based analytics, this study establishes a severe regional focus imbalance in Bulgarian media. The empirical evidence provides a stark quantitative confirmation of a "Sofia-centric" agenda, where the capital's narratives effectively dominate the national discourse, leaving most the country's administrative districts in a state of relative invisibility. This concentration of media attention is not merely a geographic artifact but a structural symptom of a captured media environment where national political theater frequently overshadows localized regional needs. The findings highlight how this imbalance contributes to the creation of "information deserts" in peripheral regions. In these areas, the news cycle is rarely driven by local development or social progress. Instead, these districts typically gain visibility only when associated with negative events or systemic crises. The high frequency of crisis-related sentiment – dominated by geopolitical tensions and national political transitions – demonstrates a media ecosystem that prioritizes sensationalism and high-conflict narratives to drive engagement. This trend highlights the need for a more decentralized reporting strategy to support sustainable development and civic engagement across all Bulgarian districts. While this study provides a comprehensive overview of current trends, it also serves as a foundation for future inquiries. Subsequent research could expand the scope of the media intelligence pipeline to include social media platforms and video-sharing sites, enabling a multimodal analysis of how regional narratives are visually and verbally constructed. Such efforts are crucial for promoting a more pluralistic and equitable information landscape.

Acknowledgment

This work was financially supported by the UNWE Research Programme (Research Grant No. 22/2024/A).

References

1. Indzhov, I. (2024). Bulgarian Media Since 1989: From Instrumentalization to Capture. Central European Journal of Communication, 17(36), 185-203.
2. Lozanov, G., Spassov, O., & Popova, K. (2011). Media and politics. Foundation Media Democracy Konrad-Adenauer-Stiftung: Sofia.
3. Pancheva, Y. (2025). Visual journalism in Bulgarian regional online media–meta-analysis. In Media Environment and Media Research-Methodologies, Meta-analysis, Trends (pp. 93-106).
4. Antonova, V., Georgiev, A., Dragomir, M., Thompson, M., Walter, M., Chan, Y. Y., ... & Tambini, D. (2013). Mapping digital media: Bulgaria. A report by Open Society Foundations.
5. Mitrović, M., Vujović, M., & Stamenković, I. (2025). Coverage of the European parliament elections by media portals in Serbia. Media and Communication, 13.
6. Kiely, K. P., & Gargova, S. (2024). A Framework for analysing disinformation narratives: Ukrainian Refugees in Bulgaria. Cybernetics and Information Technologies, 24(4), 191-209.
7. Lozanova, D., Voynova, S., Gabova, S., & Lomeva, S. (2017). Media Content Analysis on Online Hate Speech. National report Bulgaria, Positive Messengers, Sofia.
8. Nedelcheva, S. (2017). Crossing borders on the Balkan route: representation of migration in online news. Studies in Linguistics, Culture, and FLT, 2(1), 53-70.
9. Ivanova, I. (2025). Conflicting narratives: How the media creates positive and negative portraits of civil society in Bulgaria. In Media Environment and Media Research-Methodologies, Meta-analysis, Trends (pp. 107-119).
10. Hristova, G. (2021). Text analytics in Bulgarian: An overview and future directions. Cybernetics and Information Technologies, 21(3), 3-23.

11. Dinkov, Y., Koychev, I., & Nakov, P. (2019). Detecting toxicity in news articles: Application to Bulgarian. In Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP-2019) (pp. 247-258).
12. Nickl, P., Moussaïd, M., & Lorenz-Spreen, P. (2025). The evolution of online news headlines. Humanities and Social Sciences Communications, 12(1), 1-13.
13. Belev, I. (2025). Use cases for integrating Enterprise Content Management systems with Business Process Management systems. In Innovative Information Technologies for Economy Digitalization (IITED-2025) (pp. 95-99).
14. Stefanov, G. (2025). Data Types in Business Process Management: classification, comparison, and analytical potential. In Innovative Information Technologies for Economy Digitalization (IITED-2025) (pp. 257-261).
15. Tabov, Y. (2025). Technological Features of Modern Video-Sharing Platforms. In Innovative Information Technologies for Economy Digitalization (IITED-2025) (pp. 145-150).
16. Marzovanova, M., & Mihova, V. (2025). Business logic unit testing optimization in. NET using AI. In Innovative Information Technologies for Economy Digitalization (IITED-2025) (pp. 250-256).
17. Chernyshova, G., Taran, E., Firsova, A., & Vavilina, A. (2025). Monitoring of Sustainable Development Trends: Text Mining in Regional Media. Sustainability, 17(7), 3122.
18. Staiculescu, O. (2025). Digital Voices of Sustainability: A Comparative Analysis of Public Perceptions in Romania and Bulgaria. Annals-Economy Series, 5, 30-36.

Methodological Dimensions in Structuring Content on Video Platforms

Yavor Tabov

University of National and World Economy, Sofia, Bulgaria
jtabov@unwe.bg

Abstract. This paper examines the evolution and functionality of web-based video platforms such as YouTube, TikTok, and Netflix. These platforms support large-scale content creation, distribution, and user interaction. They rely on mechanisms for structuring and organizing video content, including metadata and recommendation algorithms, to deliver personalized videos and shape user engagement and marketing strategies. The paper explores the role of descriptive, structural, administrative, and media-specific metadata, often implemented through HTML tags, in improving searchability, accessibility, streaming, and overall user experience. By transforming raw video into organized and usable content, metadata ensures effective platform functionality and enhances audience interaction. In conclusion, this paper presents findings derived from this study.

Keywords. Video platforms, Web-based applications, Recommendation systems, Metadata, Html tags.

1. Nature and Trends of Video Platforms

Video platforms are dynamic digital environments that support large-scale content creation, distribution, and interaction. Platforms such as YouTube, TikTok, and Twitch have expanded rapidly, attracting millions of users and becoming central spaces for communication, entertainment, and information exchange. Their nature is defined not only by video hosting but also by strong social and interactive features [3].

Platforms such as Netflix and Disney+ are experiencing rapid growth, with forecasts projecting over 1.5 billion subscribers and 18.2% market penetration by 2025. This surge reflects the rise of connected viewing, which reshapes how screen media is created, distributed, and consumed. Streaming services not only transform the economic structures of traditional television broadcasting but also fundamentally change media consumption patterns. Viewers now experience on-demand, personalized, and multi-device content, which contrasts with the fixed schedules and linear programming of conventional television. This shift highlights the transformative impact of streaming platforms on audience behavior and the broader media ecosystem [4].

In general, video platforms have rapidly evolved into major global social and media ecosystems, each with distinct characteristics, user bases, and research interests. Platforms such as YouTube, launched in 2005 and acquired by Google in 2008, have become the largest and most influential, attracting billions of views and millions of content creators worldwide. These systems combine video hosting with strong social and interactive features, shaping communication, entertainment, and information exchange across diverse audiences. Alongside YouTube, other video platforms focus on specialized content and real-time interaction, exemplified by Twitch.

Twitch, launched in 2011, is a live-streaming video platform that had 140 million monthly active users in 2022. Its content mainly consists of gamers streaming live video and audio, accompanied by interactive text-based chat. Research on Twitch emphasizes live streaming techniques, real-time interaction, and community dynamics, highlighting the social and participatory aspects of the platform [3].

Streaming platforms offer a more interactive viewing experience, giving users greater control over what content they watch, when they watch it (time-shifting), and where they watch it (place-shifting). Unlike traditional TV, audiovisual consumption is no longer bound by fixed schedules. Platform features that enhance interactivity provide users with a sense of control over their decisions and actions, improving both the usefulness and appeal of the service [4].

Short-form video platforms such as TikTok, Instagram Reels, and YouTube Shorts have introduced new trends that are reshaping global marketing strategies. These platforms are characterized by algorithm-driven content feeds, rapid content creation, and highly interactive user engagement. Research highlights how these features influence marketing effectiveness, as the design of recommendation algorithms, patterns of content production, and user interaction dynamics determine which videos gain visibility and how audiences respond. Understanding these unique characteristics is essential for developing effective marketing approaches in the short-form video ecosystem [5]. Central to this ecosystem are the recommendation systems, advanced AI mechanisms that analyze user behavior – including watch time, engagement, and content history – to curate personalized content feeds.

Recommendation systems on short-form video platforms are advanced AI mechanisms that analyze user behavior, including watch time, engagement (likes, comments, shares), and content history. These algorithms curate personalized content feeds, ensuring high audience retention by delivering videos aligned with individual user preferences. While this enables marketers to reach highly relevant audiences quickly, it also creates challenges, as marketing content must compete with vast amounts of user-generated material. Successful campaigns leverage the algorithm by producing engaging, shareable content that can go viral, but effectiveness depends on carefully aligning marketing strategies with the platform's algorithmic logic and user behavior [9].

Video content applications can be understood through four key perspectives that define their nature and functionality. From a content perspective, these platforms primarily enable users to create and broadcast user-generated videos, either through uploads or live streaming, making video the central medium of communication. From the user perspective, content creators share videos that are personally meaningful to engage audiences, with some professionalizing their work and becoming micro-celebrities who generate income through revenue sharing, affiliate marketing, or gifts from viewers. From a social perspective, interactive features such as liking, sharing, subscribing, commenting, chatting, and donations foster the formation of virtual communities and enhance social engagement around video content. From a service perspective, ranking, recommendation, and personalization algorithms are employed to improve content delivery and optimize user experience. Overall, these four perspectives highlight that video platforms are complex socio-technical ecosystems where content, users, social interaction, and algorithmic services interact to shape platform functionality and user engagement.

Overall, these four perspectives highlight that video platforms are complex socio-technical ecosystems where content, users, social interaction, and algorithmic services interact to shape the platform's functionality and user engagement [3].

2. Metadata as a Structural and Methodological Element in Video Content

Video platforms are environments in which large volumes of video content are continuously created, stored, and reused. A key characteristic of these platforms is the need to organize and describe video materials in a way that allows efficient access and retrieval. In practice, this often requires human involvement in reviewing video files and assigning descriptive information. Such activities include content moderation, tagging, cataloging for educational or cultural collections, and preparing videos for search and recommendation systems.

An example of manual video archiving illustrates how archivists review video materials and create metadata, which represents a key component of structured data used to describe, organize, and contextualize digital content, capturing information about the people, locations, and actions shown in the footage. This practice highlights a fundamental aspect of video platforms during manual archiving: structured descriptions are essential to transform raw video files into organized, searchable, and usable content, enabling effective management and retrieval across different applications and audiences. [9].

Metadata is often perceived as abstract or complex concepts, partly because it is commonly associated with artificial intelligence and advanced algorithmic systems. It represents simple, descriptive information that has long been used in everyday contexts. Traditional examples, such as library catalog records, show that metadata plays a fundamental role in organizing, locating, and managing information, making it an essential and practical component of information systems [11].

Metadata is increasingly used in web applications to support content organization, navigation, and search engine optimization. As descriptive information about digital content, metadata helps users, platforms, and search engines better understand and access web materials, especially media-rich content such as videos and images. In practice, metadata is commonly implemented through HTML elements such as <meta>, <title>, and <link> tags, which provide structured information about web pages. However, metadata usage must be balanced, since excessive or poorly structured metadata can negatively affect web performance. Factors such as page load time, server response, and system throughput may be influenced by how metadata is applied [10].

Video platforms host rapidly growing volumes of diverse video content, which requires effective organization and classification to support search, access, and user experience. Video classification is essential for both usability and platform monetization, but it faces significant challenges due to content diversity, variable quality, and insufficient metadata. Although many classification approaches rely on visual analysis, methods that depend on descriptive information remain limited because creators often provide incomplete descriptions. As a result, poor metadata reduces content discoverability and complicates user navigation on video platforms. [1].

Extracting information from video content and organizing it through metadata is a key process in video platforms. Metadata such as title, description, creator, date, duration, and tags enables videos to be stored, accessed, and classified in a structured way. This information can be generated manually or automatically using technologies such as machine learning and computer vision. Video classification, based on content characteristics or predefined criteria, supports efficient video management and improves user experience by making content easier to find and organize [8].

Figure 1 illustrates the key roles of metadata in video content.

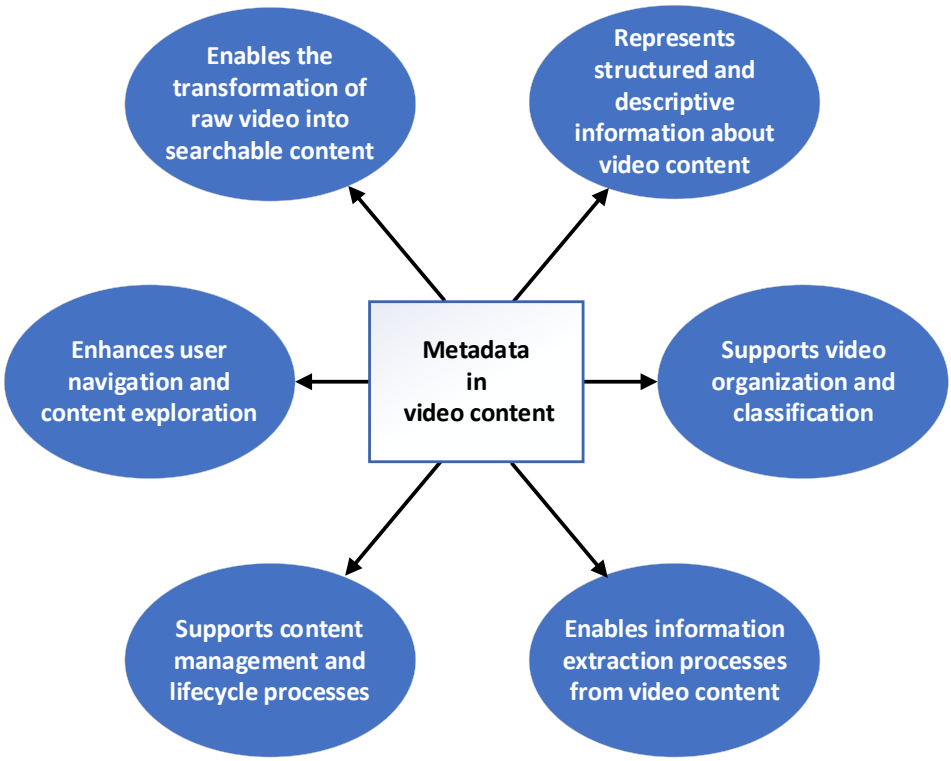


Fig. 1. Roles of metadata in video content

In conclusion, metadata plays a crucial role in the functioning of video platforms by enabling the organization, discovery, and efficient delivery of video content. In the context of modern video platforms, metadata is essential for transforming raw video files into accessible and meaningful content, thereby enhancing user experience, content management, and overall platform performance.

3. Metadata Usage in Web-Based Video Platforms

Popular video platforms such as YouTube, Twitch, Instagram, TikTok, and Netflix function as web-based applications built on standard web technologies. As such systems, they rely on structured metadata to describe, manage, and deliver video content efficiently. Metadata plays a key role in shaping the methodological dimensions of content structuring, guiding effective content discovery, streaming optimization, and user interaction by determining how video resources are represented and processed within web-based environments. Structuring content on these platforms relies heavily on different types of metadata, which define how content is described, organized, accessed, and delivered. In web-based applications, which are implemented and presented through HTML, different tags and meta tags are used to represent and structure metadata. According to the scientific literature, metadata in the context can be categorized into the following types: descriptive, structural, administrative and media-specific.

Descriptive metadata provides information about the content and purpose of a web page, helping search engines and users understand what the page is about, and it plays a central

role in content discovery and search engine optimization. Here, the following HTML elements are typically included:

• Title tags specify the page title displayed in search engine results and browser tabs, directly influencing click-through rates. HTML titles have always been and remain the most important HTML signal that search engines use to understand what a page is about. Bad titles on web pages are like having bad book titles in the examples above. In fact, if HTML titles are deemed bad or not descriptive, search engine may change them [7].

• Meta description tags provide a concise overview of a page, helping users determine its relevance to their search [10]. This information also influences user engagement and click-through rates, although search engines may modify it based on query relevance [7].

• Keywords are metadata tags containing terms related to the page content; although their importance has decreased, they still provide contextual signals [10].

Structural metadata defines how different parts of a resource are organized and related. It includes elements such as unique identifiers, page numbers, and structural features like tables of contents and indexes. In addition, it describes relationships between datasets and supports linking between individual components of a digital resource [6]. This category typically involves the following HTML elements:

• Sitemap metadata helps search engines crawl and index all pages of a website systematically, providing information about page hierarchy, update frequency, and relative priority, which enhances discoverability, ensures important content is indexed, and supports overall site optimization.

• Link tags define relationships between documents and external resources, such as stylesheets, alternate versions, or canonical URLs [10].

Administrative metadata describes the permissions, rights, preservation, and usage of data, supporting proper content management, compliance with legal and licensing requirements, and long-term accessibility within web-based applications [2]. Common HTML elements used to represent this metadata include:

• HTTP header metadata defines rules related to content type, caching behavior, and security policies.

• Author and copyright tags provide information about content ownership, authorship, and licensing conditions.

Media-specific metadata refers to descriptive information associated with rich media elements such as images, videos, and PDF files, which are widely used in modern web applications, improving visibility, sharing, and accessibility of media content. The following HTML elements and attributes are often employed for media metadata:

• Open Graph (OG) tags control how media content appears when shared on social platforms.

• PDF metadata supports the organization and linking of document-based resources.

• Alt text for images enhances accessibility for users with visual impairments and improves image indexing for search engines [10].

Fig. 2 presents examples of metadata tags corresponding to the four categories in the context of web-based video platforms.

Metadata is crucial for web-based video platforms, as it enables efficient organization, discovery, and delivery of content. By using descriptive, structural, administrative, and media-specific metadata, platforms can improve searchability, optimize streaming, enhance accessibility, and support user engagement, while ensuring proper management, ownership, and technical performance of video resources.

Metadata tags

```
Descriptive Metadata
<meta name="title" content="Sample Video Title">
<meta name="description" content="A platform for streaming and discovering video content">
<meta name="keywords" content="video, streaming, clips, live, entertainment">
<meta name="subject" content="Video streaming and media content">

Structural Metadata
<link rel="canonical" href="https://videoplatform.com/watch/12345">
<link rel="alternate" href="https://videoplatform.com/mobile/watch/12345" media="only
screen and (max-width: 640px)">
<link rel="sitemap" href="https://videoplatform.com/sitemap.xml">

Administrative Metadata
<meta name="author" content="VideoPlatform Inc.">
<meta name="copyright" content="VideoPlatform Inc. 2025">
<meta http-equiv="Content-Type" content="text/html; charset=UTF-8">
<meta http-equiv="cache-control" content="max-age=3600">

Media-Specific Metadata
<meta property="og:title" content="Sample Video Title">
<meta property="og:type" content="video">
<meta property="og:url" content="https://videoplatform.com/watch/12345">
<meta property="og:image" content="https://videoplatform.com/thumbnails/12345.jpg">
<meta property="og:site_name" content="VideoPlatform">
<meta property="og:description" content="Watch this video on VideoPlatform, featuring
highlights, clips, and live content.">
<meta name="video:duration" content="PT5M30S">
<meta name="video:upload_date" content="2026-01-07">
<meta name="video:rating" content="PG">
<meta name="video:keywords" content="action, highlights, clips, trending">
<meta name="video:author" content="ContentCreator123">
```

Fig. 2 Sample metadata tags for web-based video platforms

4. Conclusion

Based on the presented research, the following conclusions can be drawn:

- Video platforms evolve into major global social and media ecosystems, characterized by diverse user groups, platform features, and research interests.
- Metadata is essential to the operation of video platforms, enabling efficient content organization, discovery, and delivery. It transforms raw video files into accessible, meaningful content, enhancing user experience, management, and platform performance.
- In web-based video platforms, implemented and rendered via HTML, various tags and meta tags are employed to represent and organize metadata.
- Through descriptive, structural, administrative, and media-specific metadata, platforms enhance searchability, optimize streaming processes, improve accessibility, and foster user engagement.

Acknowledgement

This work was financially supported by the UNWE Research Programme (Research Grant No. 26/2026/Б).

References

1. Algur, S. P., & Bhat, P. (2015). Metadata based classification and analysis of large scale web videos. International Journal of Emerging Trends and Technologies in Computer Science, 4(3).

2. Allen, R. B. (2020). Metadata for social science datasets.
3. Bartolome, A., & Niu, S. (2023, April). A literature review of video-sharing platform research in HCI. In Proceedings of the 2023 CHI conference on human factors in computing systems (pp. 1-20).
4. Evens, T., Henderickx, A., & Conradie, P. (2024). Technological affordances of video streaming platforms: Why people prefer video streaming platforms over television. European Journal of Communication, 39(1), 3-21.
5. Huang, A., & Zainol, Z. (2025). The impact of short video platform characteristics on marketing effectiveness: A bibliographic review and future research directions. International Journal of Global Economics and Management, 6(1)
6. Mosha, N. F., & Ngulube, P. (2023). Metadata standard for continuous preservation, discovery, and reuse of research data in repositories by higher education institutions: A systematic review. Information, 14(8), 427.
7. Patel, B. V., & Gaharwar, R. D. (2018). Search engine optimization (SEO) using HTML Meta-Tags. International Journal of Scientific Research in Science and Technology (IJSRST), 4(9), 298-302.
8. Reddy, T. R., Sreekari, P., Reddy, J. N. K., & Jyothsna, V. (2023). A deep learning approach for video metadata generation and classification. International Journal of Innovative Research in Technology, 9(12),
9. Shakurov, M., Korobkin, D., Fomenkov, S., & Golovanchikov, A. (2021, October). Software for generation of video files metadata. In Journal of Physics: Conference Series (Vol. 2060, No. 1, p. 012029). IOP Publishing.
10. Sharma, A., & Gupta, V. (2024). An analysis of impact of metadata image, video and PDF in web applications using performance testing. Journal of Electrical Systems, 20(11s).
11. Weller, A., Bleisteiner, W., Hufnagel, C., & Iber, M. (2024). The Future is Meta: Metadata, Formats and Perspectives towards Interactive and Personalized AV Content. arXiv preprint arXiv:2407.19590.

A Taxonomy of Simulation - Based Evaluation for Online Machine Learning

Bozhidar Bahov

University of National and World Economy, Sofia, Bulgaria
bozhidar.bahov@unwe.com

Abstract. Online machine learning models learn incrementally from continuous data streams, where ground truth is often delayed, scarce, or entirely absent. This makes real-world evaluation during development impractical for a meaningful class of tasks. Simulation and synthetic data fill that gap, but the approaches researchers use have developed independently across subfields with no unifying framework connecting simulation type to task characteristics. We propose a taxonomy of simulation-based evaluation for online machine learning organized along two axes: whether a feedback loop exists between the model and its environment, and whether ground truth is intrinsic to the data generation process or injected into a background stream. Crossing these axes yields four paradigms: passive statistical generators, active agent-based environments, controlled synthetic injection, and active synthetic injection, each with published instances in different subfields. To illustrate the practical consequences of paradigm mismatch, we evaluate a recommendation model under both passive and active paradigms with identical task conditions and a concept drift event. Passive evaluation overestimates convergence speed by a factor of two and post-drift recovery speed by a factor of three. The taxonomy provides a structured basis for reasoning about evaluation design and makes visible a concrete tooling gap: injection-based paradigms, unlike passive and active evaluation, lack standardized domain-agnostic frameworks.

Keywords. Online Machine Learning, Simulation-based Evaluation, Concept Drift, Synthetic Data, Evaluation Methodology

1. Introduction

A supervised learning model trained on a static dataset can be evaluated straightforwardly: hold out a test set, measure performance, and report it. Online learning models cannot. They learn incrementally from a continuous data stream, updating after each observation rather than training on a fixed corpus. There is no held-out set because the data arrives once, in order, and is used immediately. There is no clean separation between training and testing.

This creates a real evaluation problem. The label for a prediction often arrives long after the prediction was made. In fraud detection, an account flagged as suspicious may not be confirmed as fraudulent until an investigation completes, days or weeks later. In clinical risk scoring, whether a patient deteriorates may not be known until after discharge. The literature formally recognizes this gap between a prediction and the arrival of its label as verification latency [1, 2, 3]. During that window, the model continues to operate without feedback, and performance degradation goes undetected. In more extreme cases like novel attack types, emerging topics, previously unseen phenomena, labels may never arrive at all, because the signal was not anticipated when the data collection system was designed. Gama et al. [1] and Ditzler et al. [2], alongside foundational work by Krempl et al. [3] on verification latency and

Souza et al. [4] on extreme verification latency, discuss how this delay makes real-world evaluation during development practically infeasible for a meaningful class of online learning tasks.

The practical consequence is that online ML research relies heavily on simulation. Developers generate synthetic data streams with known properties, run their models against them, and measure accuracy against the generator's ground truth. The SEA [5] algorithm formalized this approach early, and subsequent frameworks like MOA [6] and River [7] built mature tooling around it. For the tasks these tools were designed for (stream classification, regression, supervised anomaly detection) they work well. The model observes a stream it cannot influence, labels are available at each step, and prequential evaluation provides a clean accuracy estimate.

The problem is that online ML has expanded considerably beyond stream classification. Recommendation systems, adaptive resource allocation, online control, and detection tasks on unlabelled streams each introduce properties that passive synthetic generators were not designed to handle. Recommendation systems have a feedback loop: the items a model recommends determine what it learns next. Detection tasks on real-world streams lack the intrinsic ground truth that a synthetic generator supplies by construction. At the same time, adjacent fields like reinforcement learning, cybersecurity, and natural language processing have developed their own simulation-based evaluation methods, independently and for their own purposes. The result is a fragmented landscape where the right evaluation approach for a given task is largely determined by which subfield's conventions a researcher happens to follow, rather than by any principled analysis of the task's properties.

To the best of our knowledge, no published framework organizes these approaches into a taxonomy matched to task characteristics. This matters because the mismatch between evaluation paradigm and task type both suboptimal and producing predictably misleading results. A recommendation model evaluated on a passive stream of randomly drawn items gets uniform exploration for free, which is exactly what the model would not receive in deployment. A topic detection system evaluated on historical data with retrospective labels is scored against ground truth that was constructed with full knowledge of how topics eventually developed — the evaluation is reproducible, but the ground truth embeds hindsight that the model could not have had at detection time.

This paper makes two contributions. First, we propose a taxonomy of simulation-based evaluation for online machine learning organized along two axes: whether a feedback loop exists between the model and its environment, and whether ground truth is intrinsic to the data generation process or injected into a background stream. Crossing these axes yields four paradigms: passive statistical generators, active agent-based environments, controlled synthetic injection, and active synthetic injection; each with published instances across different subfields. The taxonomy's contribution is the unifying framework, not the discovery of any individual paradigm. Second, to ground the taxonomy in a concrete example, we present an illustrative demonstration visualizing one type of paradigm mismatch: a recommendation model evaluated under both the passive and active paradigms with identical task conditions and a concept drift event. We find that passive evaluation overestimates both convergence speed and post-drift adaptation speed. We are explicit about the scope of this demonstration: it visualizes one mismatch type. The broader claim that paradigm mismatch produces misleading results rests on the theoretical structure of the taxonomy itself, supported by synthesis of existing literature.

The remainder of the paper is organized as follows. Section 2 reviews relevant background in online machine learning, current evaluation practice, and simulation-based evaluation in adjacent fields. Section 3 defines the taxonomy formally. Section 4 presents the

illustrative demonstration. Section 5 discusses implications, the fragmentation of injection-based evaluation tooling, and directions for future work.

2. Background and Related Work

2.1. Online Learning and Concept Drift

Online learning, also called incremental learning, refers to the family of methods that update a model one observation at a time as data arrives, without storing or revisiting past examples. This stands in contrast to batch learning, where training and inference are separate phases operating on a fixed dataset. The defining constraint of online learning is that each data point is processed once, in order, and the model must be ready to make predictions at every step.

A central challenge in this setting is concept drift: the phenomenon where the joint distribution $P(x,y)$ underlying the data stream changes over time [1]. Gama et al. [1] distinguish several drift types based on the speed and structure of the change.

- Sudden drift is an abrupt transition from one distribution to another at a specific point in time.
- Gradual drift involves a period of overlap where both old and new distributions contribute to the stream.
- Incremental drift is a slow, continuous shift in the underlying concept, with no clear transition point.
- Recurring drift occurs when the stream cycles between previously seen distributions, such as seasonal patterns.

Lu et al. [8] provide a comprehensive review of detection and adaptation methods across these drift types.

Verification latency complicates evaluation in all of these settings. Even if a model detects a change in the input distribution, correcting its predictions requires labelled feedback, which may arrive long after the event that caused the drift. Krempl et al. [3] identify verification latency as one of the fundamental open challenges in stream mining: the gap between prediction time and label arrival creates a window where model quality is unobservable. In the most severe case, labelled feedback never arrives at all. Souza et al. [4] formalize this as extreme verification latency and develop methods for learning in environments where the assumption of eventual labeling breaks down entirely. Both challenges are important context for evaluation design: a synthetic evaluation environment that provides labels at every step is implicitly assuming conditions that many real deployment settings do not satisfy.

2.2. Current Evaluation Practice

The standard approach to evaluating online learners is prequential evaluation, introduced by Dawid [14] in 1984 and adopted as a central protocol in the stream mining literature by Bifet et al. [6] At each step, the model is first tested on the incoming example, then trained on it. This interleaved test-then-train procedure uses every example for both evaluation and learning, making efficient use of the stream without requiring a held-out set. Performance is typically reported as a rolling or cumulative accuracy over a sliding window.

The main source of data in this setting is synthetic stream generators. The SEA generator [5] introduced the idea of a labelled stream with controllable concept drift, using a threshold function over three randomly generated features. Subsequent generators extended this to different problem geometries: the Agrawal generator uses a decision tree over attributes representing loan applications, and the Hyperplane generator produces a stream where the decision boundary rotates continuously, enabling study of gradual drift. These generators share

a common property: the ground truth label at each step is defined by the same mathematical function that generates the features, making evaluation clean and reproducible.

The MOA [6] framework and its successor River [7] standardized this evaluation methodology, providing libraries of generators, classifiers, and evaluation protocols in a single toolkit. River in particular has become a practical reference implementation for online learning research, and its prequential evaluator is the default evaluation mechanism for stream classification benchmarks.

The limitation of this paradigm is scope, not correctness. Ditzler et al. [2] note that standard stream evaluation measures predictive accuracy on fixed, passively observed streams, which is exactly what it should do for stream classification tasks. The limitation appears when the same approach is applied to tasks where the model's predictions influence which data it encounters next. In those settings, the passive stream gives the model information it would not have in deployment, and the evaluation overestimates real-world performance. This is not a flaw in MOA or River but a consequence of applying tools outside the task class they were designed for.

2.3. Interactive and Agent-Based Evaluation

Reinforcement learning provides the most developed framework for evaluating models that act on their environment. The agent-environment loop, formalized in Sutton and Barto [9], defines a model (agent) that observes states, selects actions, and receives rewards, with the environment transitioning to a new state as a consequence of each action. This feedback loop is the defining property: the agent's policy determines which states it visits and therefore which training signal it receives.

OpenAI Gym [10], and its maintained successor Gymnasium, operationalized this framework as a standardized interface for building and sharing evaluation environments. A Gym environment exposes a small API (reset, step, observation, reward) that any compliant agent can interact with, making it straightforward to compare algorithms across a shared set of tasks. This infrastructure has no direct equivalent in the online learning literature.

Li et al. [11] connected this framework to recommendation by formalizing news article recommendation as a contextual bandit problem. In their formulation, the model selects which article to show a user at each step, and the click or skip it receives is the reward. The key observation is that the training data the model sees is entirely determined by its own choices: a model that never recommends finance articles never learns anything about who clicks on finance articles. This is structurally identical to the agent-environment loop in RL, and it makes passive stream evaluation inappropriate for recommendation tasks.

Despite this conceptual overlap, cross-pollination between RL evaluation practice and online learning evaluation practice has been limited. Stream mining researchers have largely continued to use passive generators even for tasks with feedback loops, while the RL community has built increasingly sophisticated simulation environments without much engagement with the concept drift and verification latency problems that stream mining has studied in depth. The taxonomy in Section 3 provides a framework for understanding both traditions as instances of a shared set of choices about simulation design.

2.4. Synthetic Data as an Evaluation Instrument

A third line of work treats synthetic data not as a stream generator with intrinsic labels, but as a set of planted signals injected into a background to test whether a model can find them. This approach addresses tasks where real-world ground truth is absent by construction: the phenomenon of interest is defined as unknown at the time it matters, so retrospective labeling cannot serve as evaluation ground truth.

Shi et al. [12] applied this principle to topic modeling evaluation, constructing synthetic corpora by embedding articles about a fictitious topic into a real news corpus. Because the planted topic was designed by the evaluators, its properties like semantic coherence, article count, and injection timing were fully specified and controllable. Evaluation then measured whether a topic model correctly identified the planted cluster against the unlabelled background.

Rosso et al. [13] implemented the same principle for cybersecurity evaluation with SAIBERSOC, injecting synthetic attacks into live Security Operation Centre infrastructure. They use the "fire drill" metaphor explicitly: a planned, controlled exercise that tests detection capability under realistic conditions without waiting for a real incident. The key advantage over historical evaluation is that the ground truth is known (the timing and nature of the injected attack are specified by the evaluators) while the background is real operational data, providing realistic noise and false positive conditions.

These two implementations were developed independently, for different domains, with no common framing connecting them. CybORG [15] extends the cybersecurity case into active territory: synthetic attacker agents (red teams) are injected into a simulated network, and defender agents (blue teams) must detect and respond while their own actions change the network state. This combines injected ground truth with an active feedback loop, making it an instance of a qualitatively different evaluation paradigm than SAIBERSOC or Shi et al., despite the surface similarity.

The shared principle across these examples - plant a known signal, measure whether the model finds it, and use the background as a realistic noise source, has been arrived at separately by researchers in NLP, cybersecurity, and network science. Section 3 unifies them under a common framework.

3. A Taxonomy of Simulation-Based Evaluation

This section proposes a taxonomy that organizes simulation-based evaluation approaches for online machine learning into four paradigms. The taxonomy is structured along two axes, each capturing a property that determines which evaluation approach fits a given task. All four paradigms have published instances across different subfields; the contribution here is the unifying framework, not the identification of any individual paradigm.

3.1. Distinguishing Axes

The first axis asks whether a *feedback loop* exists between the model and its data source. In some online learning tasks, the model passively receives data from an external process it cannot influence. A classifier processing sensor reading, for example, has no effect on what the sensors report next. In other tasks, the model's outputs directly shape what data it encounters in the future. A recommendation system that shows a user an article about finance will observe a click or a skip on that specific article, not on the sports article it chose not to show. The training signal the model receives is a consequence of its own decision.

This distinction matters for evaluation because simulation environments that ignore the feedback loop give the model access to information it would not have in deployment. A model evaluated on a fixed stream of uniformly sampled items learns from all of them equally. The same model deployed in production only learns from items it chose to present, creating gaps in its knowledge that the fixed-stream evaluation never revealed.

The second axis concerns the *origin of ground truth*. We distinguish between intrinsic and injected ground truth.

Intrinsic ground truth is native to the data generation process. When a synthetic stream generator like SEA [5] produces a data point, the label is defined by the same mathematical

function that generates the features. In an interactive environment, the reward signal is defined by the environment's own rules. In both cases, ground truth exists because the simulation creates it as part of producing the data.

Injected ground truth works differently. Synthetic entities with known labels are inserted into a background stream or environment that may lack labels of its own. A cybersecurity researcher might inject simulated attacks into a live network feed to test whether a detection system finds them [13]. A natural language processing researcher might plant articles about a fictitious topic into a news corpus to test whether a topic model identifies the new cluster [12]. In these cases, the background data carries no reliable ground truth (real attacks and real emerging topics are unlabelled by definition). The evaluation measures whether the model identifies the injected entities against this unlabelled background.

The distinction between intrinsic and injected ground truth is sometimes confused with the distinction between synthetic and real data. The SEA generator, for instance, is entirely synthetic, yet its ground truth is intrinsic: the labels are part of the generation process. The key question is not whether the data is real or synthetic, but whether ground truth is a native product of stream creation or an artificial overlay applied to a separate background.

Crossing these two axes produces four paradigms, summarized in Table 1.

Table 1. Four paradigms of simulation-based evaluation for online machine learning.

Paradigm	Feedback Loop	Ground Truth
Passive Statistical Generator	No	Intrinsic
Active Agent-Based Environment	Yes	Intrinsic
Controlled Synthetic Injection	No	Injected
Active Synthetic Injection	Yes	Injected

The following subsections define each paradigm formally and identify the class of tasks for which it is appropriate.

3.2. Paradigm 1: Passive Statistical Generators

A passive statistical generator produces a temporally ordered stream $S=((x_1,y_1),(x_2,y_2),\ldots,(x_T,y_T))$ of labelled observations. At each time step t the model receives an input x_t, produces a prediction $\hat{y}_t = f(x_t;\theta_t)$, is evaluated against the true label y_t via a loss function $L(y_t, \hat{y}_t)$, and updates its parameters: $\theta_{t+1} = update(\theta_t, x_t, y_t)$

The defining property of this paradigm is independence: the generating distribution $P(x_t, y_t)$ does not depend on any of the model's prior predictions $\{\hat{y}_1,\ldots,\hat{y}_{t-1}\}$. The stream unfolds identically regardless of what the model does with it. Ground truth is intrinsic, defined by the mathematical function that generates the data (a hyperplane for SEA [5], a set of classification rules for Agrawal, and so on).

This is the dominant evaluation paradigm in the online learning literature. The SEA algorithm [5] introduced the concept of a streaming ensemble evaluated against a synthetic generator with abrupt concept drift. The MOA [6] framework and its successor River [7] standardized this approach, providing libraries of generators with controllable drift characteristics and prequential (interleaved test-then-train) evaluation protocols [6].

Passive generators are appropriate for tasks where the model genuinely has no influence on the data stream: stream classification, regression on sensor data, supervised anomaly detection where explicit labels arrive continuously. They are well understood, reproducible, and supported by mature tooling.

Their limitation is specific. When applied to tasks that involve a feedback loop between the model and its environment (recommendation, control, active resource allocation), passive generators supply the model with exploration it would not receive in deployment. Section 3.6 discusses this failure mode, and Section 4 visualizes it through a concrete example.

3.3. Paradigm 2: Active Agent-Based Environments

An active agent-based environment introduces a stateful simulation in which the model's outputs influence its future inputs. Formally, the environment maintains a state s_t. At each time step t_t, the model observes a context x_t derived from s_t, selects an action $a_t = \pi(x_t;\theta_t)$ according to its policy π, and receives a reward $r_t = R(s_t,a_t)$ from the environment. The environment then transitions to a new state $s_{t+1} = T(s_t, a_t)$, and the model updates its parameters: $\theta_{t+1} = update(\theta_t, x_t, a_t, r_t)$.

The critical difference from Paradigm 1 is that the marginal distribution of observations $P(x_t)$ becomes a function of the model's policy π. The environment's transition dynamics T remain fixed, but because the policy determines which actions are taken and therefore which states are visited, the actual data distribution the model encounters is entirely policy-dependent. The model's choices shape the data it will learn from. Ground truth remains intrinsic; the reward function R is native to the environment definition.

This formulation is standard in the reinforcement learning literature, where the agent-environment loop is foundational [9]. OpenAI Gym [10] and its successor Gymnasium established standardized interfaces for building and sharing such environments. Li et al. [11] formalized news article recommendation as a contextual bandit problem, demonstrating how the feedback loop between recommendation decisions and user responses creates a fundamentally different evaluation challenge than passive prediction.

Active environments are appropriate for tasks where the model's predictions determine what happens next. This includes recommendation systems, control tasks, dialogue systems, and any setting where the model acts on its environment rather than simply observing it.

The limitations of this paradigm are practical rather than conceptual. Building a realistic active environment requires substantially more engineering effort than configuring a passive generator. The reward function and state transition model are simplifications of real-world dynamics, and the gap between simulated and actual behaviour (sometimes called the simulation-to-reality gap) means that strong performance in simulation does not guarantee deployment success.

3.4. Paradigm 3: Controlled Synthetic Injection

Controlled synthetic injection addresses a problem that neither passive generators nor active environments can solve: evaluation when ground truth does not naturally exist. Consider novel topic detection in a news stream. At the moment a new topic begins to emerge, no one has labelled it as such. Detection must happen before ground truth is available. The same applies to zero-day cyberattacks, emerging health threats, and other phenomena whose defining characteristic is that they are unknown at the time they matter most.

In this paradigm, a background stream $B = (b_1, b_2, \ldots, b_M)$ represents a temporally ordered sequence of real or realistic data, typically unlabelled. A set of synthetic items $S = (s_1, s_2, \ldots, s_K)$ with known labels is injected into B according to a controlled schedule that maps each synthetic item to a position in the timeline, producing a combined temporal sequence $C = (c_1, c_2, \ldots, c_T)$ where $T = M + K$. The model processes C in order without knowledge of which items are injected.

Evaluation requires careful treatment of both synthetic and background items. For any item c_t in the combined stream, the model produces a prediction p_t. If c_t originates from the

synthetic sequence *S*, its ground truth label is known from the injection design. If c_t originates from the background sequence *B*, its true status is unverified: a real-world background stream may contain organic anomalies, genuine emerging topics, or authentic attacks that were never labelled. The evaluation framework must account for this. Predictions on injected items measure whether the model detects the planted signals (recall). Predictions on background items that the model flags as anomalous require secondary analysis: they may be false alarms, or they may be legitimate discoveries that the background stream's lack of labelling cannot distinguish. Reporting both detection performance on injected items and the rate of flags on background items gives a complete picture of model behaviour.

Rosso et al. [13] implemented this paradigm for cybersecurity evaluation with SAIBERSOC, which injects synthetic attacks into operational Security Operation Centre infrastructure to measure detection accuracy and response times. They describe the approach using the "fire drill" metaphor: a planned, controlled test of detection capability under realistic conditions. Shi et al. [12] applied the same principle to natural language processing, constructing synthetic corpora with planted topic structure to evaluate topic modelling algorithms. These implementations were developed independently for their respective domains.

This paradigm is appropriate for detection and discovery tasks on streams where ground truth is absent or unreliable: unsupervised anomaly detection, novel topic identification, emerging threat recognition, and first story detection. Its primary limitation is that synthetic content may not perfectly replicate the statistical properties of organic emergence. A well-designed injection can control for difficulty (varying semantic distance, signal density, or timing), but the realism of the planted signals depends on domain expertise in designing them. There is also no standardized, domain-agnostic toolkit for this paradigm comparable to MOA or Gym; existing implementations are fragmented across application-specific projects.

3.5. Paradigm 4: Active Synthetic Injection

The fourth paradigm combines the feedback loop of active environments with the injected ground truth of controlled injection. Synthetic entities with their own labels are introduced into an active environment where the model's actions influence future states. The model must identify and respond to the injected entities while its own decisions alter the conditions under which detection occurs.

A concrete instance is adversarial evaluation in simulated network defense. In environments like CybORG [15], synthetic attacker agents (red teams) with known attack patterns are injected into a simulated network. Defender agents (blue teams) must detect and mitigate these attacks in real time. The defender's actions (isolating hosts, blocking traffic, deploying countermeasures) change the network state, which in turn affects both the attacker's progression and the defender's future observations. Ground truth is injected (the synthetic attacker's presence and behaviour are known to the evaluator) and the feedback loop is active (the defender's responses shape the environment).

This paradigm should be distinguished from standard adversarial robustness testing, where input perturbations (noise added to images, for example) test a model's sensitivity to corrupted inputs. In that setting, the ground truth label does not change: a perturbed image of a cat is still a cat. The perturbation tests input robustness within Paradigm 2. Paradigm 4 specifically involves the injection of new entities that create new ground truth labels, such as the presence of an attacker that did not previously exist in the environment.

This paper characterizes Paradigm 4 conceptually and identifies existing instances. The illustrative demonstration in Section 4 focuses on the boundary between Paradigms 1 and 2.

3.6. Paradigm Mismatch

The taxonomy enables identification of predictable failure modes that arise when a task is evaluated under the wrong paradigm. These failure modes are theoretical deductions from the structure of the taxonomy, supported by synthesis of the literature reviewed in Section 2. We describe two types.

- **Mismatch Type A** occurs when a task with an active feedback loop is evaluated using a passive generator. A recommendation model, for example, evaluated on a fixed stream of uniformly sampled items, receives optimal exploration for free. Every item appears with equal frequency, so the model learns about all options regardless of its own preferences. In deployment, the model's policy concentrates its experience on items it already favors, starving it of information about alternatives. The passive evaluation overestimates both baseline performance and the model's ability to adapt after concept drift. Section 4 visualizes this mismatch through a controlled comparison.
- **Mismatch Type B** occurs when a detection task requiring injected ground truth is evaluated using only historical data with retrospective labels. A topic detection system evaluated on a historical news corpus, for instance, might report strong performance because topics were labelled after the fact, with full knowledge of how they developed. The evaluation is reproducible (the dataset and labels are fixed), but the ground truth itself is confounded by hindsight bias. Retrospective labeling makes it impossible to establish a precise onset point for when a signal first appeared in the stream, which prevents accurate measurement of detection latency. Controlled synthetic injection (Paradigm 3) resolves this by planting signals with known properties at known times. This mismatch is a conceptual argument; we do not visualize it empirically in this paper.

Other mismatch configurations are possible (using injection-based evaluation for a task that has intrinsic ground truth, for example), but these two represent the most commonly encountered cases in practice.

4. Illustrative Demonstration

To ground the taxonomy in a concrete example, this section presents a controlled comparison visualizing Mismatch Type A from Section 3.6: a recommendation model evaluated under both the passive and active paradigms with identical task conditions and a single concept drift event. The goal is not to establish a general empirical claim about the magnitude of this effect, but to show what the mismatch looks like when everything except the item selection mechanism is held fixed.

4.1. Experimental Setup

The task is context-dependent item recommendation. The environment contains two users (Tom and Anna), two time-of-day contexts (morning and afternoon), and seven candidate items (politics, sports, music, food, finance, health, camping). An oracle assigns one preferred item to each of the four (user, context) pairs. At step 2000, all four preferences change simultaneously. The model receives no signal that a drift has occurred.

The model is "*ContextualFunkMF*", a matrix factorisation model built on River's "*FunkMF*" [7] extended with a composite user-context key. At each step t, the model receives a user u_t and a context vector x_t drawn by an independent random process, selects or observes an item i_t, receives a binary reward $r_t \in \{0, 1\}$ reflecting whether i_t matches the oracle's current preference, updates its parameters, and is then evaluated by asking whether its current greedy ranking would produce the correct item. Evaluation is identical across both conditions: rolling accuracy over a window of $w = 100$ steps.

The two conditions differ only in how i_t is determined.

Passive condition. The item is drawn uniformly at random from the candidate set I:

$$i_t \sim \text{Uniform}(I)$$

The model has no influence over which item it observes. The training signal covers all items roughly equally regardless of what the model currently believes about them.

Active condition. The item is selected by the model via an epsilon-greedy policy with $\varepsilon = 0.15$:

$$i_t = \begin{cases} \text{Uniform}(I) & \text{with probability } \varepsilon \\ \arg\max_{i \in I} f(x_t, i; \theta_t) & \text{with probability } 1 - \varepsilon \end{cases}$$

With probability $1-\varepsilon$, the model exploits its current beliefs and selects the item it already ranks highest. With probability ε, it explores uniformly. The training signal the model receives is therefore a consequence of its own policy: items it ranks poorly are rarely selected, and the model accumulates little information about them.

The user-context random sequence is generated by the same seed in both conditions, so Tom and Anna appear in the same order, facing the same contexts, at every step. The only thing that differs is the item each condition hands to the model to learn from.

Full code for both conditions is available at:
https://github.com/bbahov/simulation-based-evaluation-oml

4.2. Results

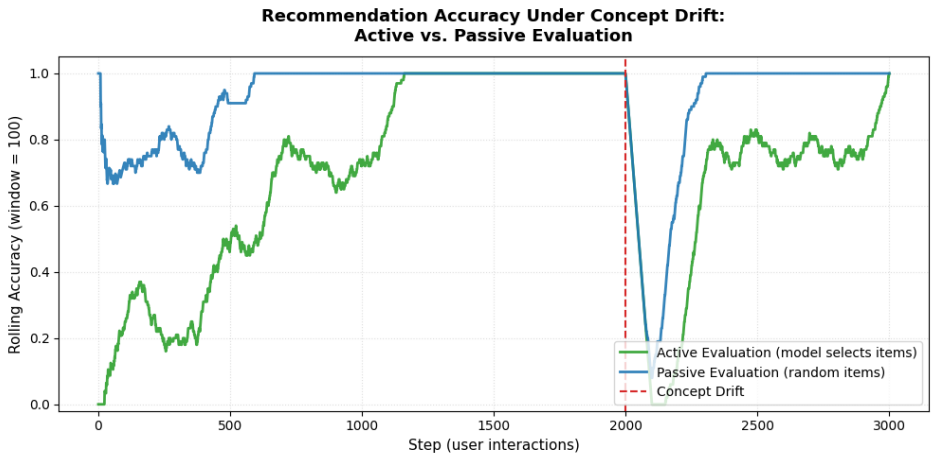


Fig. 1. shows rolling accuracy over 3000 steps for both conditions.

Pre-drift period (steps 1–2000). The passive condition reaches 1.0 accuracy by approximately step 600 and holds there for the remainder of the pre-drift period. With seven items and uniform random sampling, every item receives roughly equal exposure, allowing the model to identify the correct preference for each (user, context) pair without having to decide where to concentrate its learning. The active condition converges more erratically. The green line fluctuates sharply through the first 1000 steps and only reaches perfect accuracy at

approximately step 1200, twice as late as the passive condition. With $\varepsilon = 0.15$, the epsilon-greedy policy concentrates 85% of observations on whichever items the model currently ranks highest, leaving disfavored items chronically undersampled. The model can still learn the correct policy, but it has to find it through a much narrower exploration window.

Both conditions ultimately converge to 1.0 before drift. The difference is in the path, not the destination. The passive condition reaches the correct policy in roughly half the steps, because uniform sampling guarantees that every (user, context, item) combination receives training signal at a predictable rate. The active condition must stumble into the correct items through a combination of greedy exploitation and the limited 15% exploration budget. A practitioner comparing convergence speed across algorithms using passive evaluation would see a consistently faster learner than any of those algorithms would produce in deployment.

Post-drift period (steps 2001–3000). At step 2000, all four (user, context) preferences change. Both conditions drop sharply. The passive condition recovers to 1.0 within approximately 300 steps. The active condition spends roughly 900 steps fluctuating between 0.70 and 0.80 before spiking back to 1.0 at the end of the simulation window.

The recovery asymmetry is the clearer of the two results. Under the passive condition, the model encounters newly preferred items at the same uniform rate it always has, so drift is corrected as soon as enough observations accumulate. Under the active condition, the model enters the post-drift period with strong prior beliefs about which items to recommend. Those beliefs are now wrong. Correcting them requires the 15% exploration budget to surface the new preferred items often enough to overcome the momentum of the old policy. That process takes roughly three times as long as passive recovery. A practitioner using passive evaluation would estimate post-drift adaptation at around 300 steps. The active evaluation shows the real figure is closer to 1000.

4.3. Limitations of This Demonstration

This demonstration is deliberately small. Two users, seven items, and four preference slots are enough to make the effect visible; they are not enough to draw quantitative conclusions about the gap's magnitude in realistic systems. Several design choices constrain the scope of what the comparison shows.

The oracle is deterministic. Real users are not. A stochastic reward function would add noise to both curves, likely narrowing the visible gap at small window sizes. Whether the passive condition's ceiling would remain at 1.0 under a probabilistic oracle depends on how close the correct item's reward probability is to 1.0 compared to alternatives.

A single model architecture (*FunkMF*) and a single exploration rate ($\varepsilon = 0.15$) are used throughout. A higher exploration rate would narrow the gap by giving the active model more uniform coverage; a lower rate would widen it. The demonstration shows one point in that space, not the full relationship.

The comparison covers only the boundary between Paradigms 1 and 2. The remaining paradigm boundaries between intrinsic and injected ground truth, and between active and passive injection are argued conceptually in Section 3 and remain to be examined empirically. The claim that paradigm mismatch produces systematically misleading evaluation rests on the theoretical structure of the taxonomy, not on this comparison alone.

5. Discussion and Future Work

The taxonomy proposed in Section 3 does not argue that any one paradigm is better than another. Each fits a different class of task. Passive generators are the right tool when the model genuinely has no influence on the data it sees, and the mature tooling around them

(MOA, River) reflects years of careful engineering for exactly those tasks. The underlying issue stems from the absence of an explicit framework for deciding which tool applies to which task, rather than any inherent flaw in the tools themselves. Without that framework, the choice defaults to convention: online learning researchers reach for passive generators because that is what the literature uses, even when the task they are evaluating has a feedback loop or lacks intrinsic ground truth. Making the paradigm choice explicit, and grounding it in the two axes proposed here, is a practical improvement that does not require building anything new.

Furthermore, the taxonomy makes visible a specific gap in available tooling. For passive evaluation, River and MOA provide domain-agnostic, standardized frameworks. For active evaluation, Gym and Gymnasium serve the same function. Controlled synthetic injection (Paradigm 3) has no comparable unifying toolkit. The implementations that exist (SAIBERSOC for cybersecurity, Shi et al.'s corpora for topic modeling) were built for specific domains and are not straightforwardly portable. A researcher evaluating a novelty detection system on an e-commerce stream, for instance, would need to design an injection pipeline from scratch. Controlled synthetic injection, as characterized in this paper, remains fragmented across domain-specific implementations; building standardized, domain-agnostic tooling for Paradigm 3 is an important open problem. Addressing it for the topic detection domain specifically, through a standardized injection benchmark with controllable signal properties and known injection schedules, is one concrete direction that follows directly from this analysis. The most immediate directions for future work follow from what this paper did not do empirically. The illustrative demonstration covers one paradigm boundary (Paradigm 1 vs. Paradigm 2) on one task with one model architecture. The remaining boundaries, between intrinsic and injected ground truth and between passive and active injection, are argued conceptually here and remain to be examined through controlled comparison. Larger-scale demonstrations across multiple tasks and model families would also be needed to say anything quantitative about the magnitude of the mismatch effects described in Section 3.6.

Several limitations of this work are worth stating plainly.

1. The illustrative demonstration covers one mismatch type in a deliberately small setting; Section 4.3 describes those constraints in detail.
2. Active environments and injection pipelines introduce a simulation-to-reality gap that passive generators do not: the oracle reward function or injected signal may not accurately reflect the real-world phenomenon it is meant to represent, and a model that performs well in simulation may still fail in deployment.
3. Adopting a more appropriate paradigm is not cost-free. Building an active environment or an injection pipeline requires substantially more engineering effort than configuring a passive generator, and the taxonomy does not address that practical cost.
4. Different paradigms require different evaluation metrics: passive evaluation uses sequential accuracy, injection-based evaluation requires detection metrics (precision, recall, false positive rate, detection latency), and active evaluation uses reward-based metrics. Transitioning paradigms may mean redesigning the measurement framework, not just the simulation environment.
5. Two axes and four cells may not exhaust the relevant distinctions. Hybrid paradigms exist (a system that partially controls its own data while operating in an environment with injected signals, for instance), and additional axes may be needed to characterize them.

6. Conclusion

Online machine learning evaluation depends on simulation, but the approaches researchers use have developed separately across subfields with no shared framework connecting them. This paper proposed a taxonomy organizing simulation-based evaluation

into four paradigms along two axes: whether a feedback loop exists between the model and its environment, and whether ground truth is intrinsic to the data generation process or injected into a background stream. All four paradigms have published instances; the contribution is the unifying structure, not the discovery of any individual cell. To ground the taxonomy in a concrete example, we visualized one type of paradigm mismatch: a recommendation model evaluated under both passive and active paradigms with identical task conditions and a concept drift event. Passive evaluation overestimated convergence speed by a factor of two and post-drift recovery speed by a factor of three. The demonstration does not prove that mismatch always produces results of this magnitude, but it shows clearly what the structural difference between paradigms looks like when everything else is controlled.

The taxonomy points toward one open problem in particular. Passive and active evaluation each have mature, domain-agnostic tooling. Injection-based evaluation (Paradigms 3 and 4) does not. Existing implementations are confined to specific domains and are not straightforwardly portable. Standardized tooling for controlled synthetic injection would make this paradigm as accessible as passive generators are today, and would substantially improve the state of evaluation for detection and discovery tasks on unlabelled streams.

References

1. Gama, J., Žliobaitė, I., Bifet, A., Pechenizkiy, M., & Bouchachia, A. (2014). A survey on concept drift adaptation. ACM Computing Surveys (CSUR), 46(4), 1-37.
2. Ditzler, G., Roveri, M., Alippi, C., & Polikar, R. (2015). Learning in nonstationary environments: A survey. IEEE Computational Intelligence Magazine, 10(4), 12-25.
3. Krempl, G., Žliobaitė, I., Brzeziński, D., Hüllermeier, E., Last, M., Lemaire, V., et al., (2014). Open challenges for data stream mining research. ACM SIGKDD Explorations Newsletter, 16(1), 1–10.
4. Souza, V. M., Silva, D. F., Gama, J., & Batista, G. E. (2015). Data stream classification guided by clustering on nonstationary environments and extreme verification latency. IEEE Transactions on Neural Networks and Learning Systems, 26(4), 872–885.
5. Street, W. N., & Kim, Y. (2001). A streaming ensemble algorithm (SEA) for large-scale classification. In Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 377-382).
6. Bifet, A., Holmes, G., Kirkby, R., & Pfahringer, B. (2010). MOA: Massive online analysis. Journal of Machine Learning Research, 11, 1601-1604.
7. Montiel, J., Halford, M., Mastelini, S. M., Bolmier, G., Sourty, R., Vaysse, R., et al., (2021). River: machine learning for streaming data in Python. J. of Machine Learning Research, 22(110), 1-8.
8. Lu, J., Liu, A., Dong, F., Gu, F., Gama, J., & Zhang, G. (2018). Learning under concept drift: A review. IEEE Transactions on Knowledge and Data Engineering, 31(12), 2346-2363.
9. Sutton, R. S., & Barto, A. G. (2018). Reinforcement learning: An introduction. MIT press.
10. Brockman, G., Cheung, V., Pettersson, L., Schneider, J., Schulman, J., Tang, J., & Zaremba, W. (2016). OpenAI gym. arXiv preprint arXiv:1606.01540.
11. Li, L., Chu, W., Langford, J., & Schapire, R. E. (2010). A contextual-bandit approach to personalized news article recommendation. In Proceedings of the 19th Int. conf. on World wide web (661-670).
12. Shi, H., Gerlach, M., Diersen, I., Downey, D., & Amaral, L. A. N. (2019). A new evaluation framework for topic modeling algorithms based on synthetic corpora. In Proceedings of the 22nd International Conference on Artificial Intelligence and Statistics (pp. 816-826). PMLR.
13. Rosso, M., Campobasso, M., Gankhuyag, G., & Allodi, L. (2020). SAIBERSOC: Synthetic attack injection to benchmark and evaluate the performance of security operation centers. In Proceedings of the 36th Annual Computer Security Applications Conference (pp. 1-16).
14. Dawid, A. P. (1984). Present position and potential developments: Some personal views: Statistical theory: The prequential approach. Journal of the Royal Statistical Society: Series A (General), 147(2), 278-292.
15. Standen, M., Lucas, M., Bowman, D., Richer, T.J., Kim, J., & Marriott, D.A. (2021). CybORG: A Gym for the Development of Autonomous Cyber Agents. ArXiv, abs/2108.09118.

Social Security – Contemporary Aspects (The Education)

Konstantin Poudin

University of National and World Economy, Sofia, Bulgaria
kpoudin@unwe.bg

March ahead, o revived people,
To your future march ahead,
Forge your destiny of glory,
By the might of letters led.

March to powerful knowledge,
Let your duty be your guide!
Join the host of other peoples,
God is always by your side.

Go! For like the sun is knowledge.
On the soul it sheds its rays.
Go! A people shall not perish,
When true learning lights its days.*

Stoyan Mihaylovski
"Cyril and Methodius"

* These are the first three verses of the poem "Cyril and Methodius", written by the Bulgarian poet and public figure Stoyan Mihaylovski in the city of Ruse, Principality of Bulgaria, in 1892. Less than a decade later, the composer Panyot Pipkov created music for this poem, which became the nationally recognized Anthem of the Bulgarian Enlightenment. The English translation of the poem is taken from VESTI, Available at: https://www.vesti.bg/novini/prevodyt-na-vseuchilishtniia-himn-na-21-evropejski-ezika-e-izdaden-v-specialna-knizhka-708491

Abstract. This paper draws attention on social dimensions of security. The focus is on the education in security context, especially at individual and state level. The „Education-Security" nexus is analysed. The paper presents some of the main challenges to the quality of the educational process. The thesis is that higher quality of education, ceteris paribus, reduces insecurity, both at the state level and at the individual level.

Keywords: Security, Education**,** Education quality**.**

1. Introduction

Although states continue to wage wars and engage in conflicts as a means of protecting their interests, although societies in the countries each day are faced with various forms of violence, today security is much more than defence, internal order and justice. With the complexity of human life, security reveals itself as something much more complex and acquires broader dimensions, being a prerequisite for stability, development, prosperity and happiness.

The contemporary dimensions of security are associated with guarantee of fundamental human rights, with ecological balance and clean environment, energy diversification, both in terms of types of energy resources and their sources, economic prosperity and financial stability, development of infrastructure, including transport and communication one, countering cyber threats, and also with various social aspects - healthcare, education, social assistance, etc.

The Latin phrase "Scientia potentia est", survived to our days, shows that centuries ago people realized the power of knowledge and more specifically the power of science that creates and develops it. Knowledge as a source of power is a guarantee of security, which, as mentioned above, is associated with stability and prosperity.

Therefore, it may be assumed that education, understood as a process of acquiring knowledge, skills, competencies and at the same time upbringing, and the related educational system, within which this process takes place, are extremely important from security point of view. The result of the educational process, namely the level of education, is even of greater importance for security. The level of education is an essential qualitative characteristic of both the individual and a society.

This report, without claiming to be exhaustive, sheds light on the "Education-Security" nexus and presents the main challenges to the quality of the educational process. It presents more general views (the author's reflections) on the issue under consideration. The thesis is that higher quality of education, ceteris paribus, is a prerequisite for reducing insecurity, both at the state level and at the individual level.

2. "Education-Security" Nexus

For many people, at first glance, the „Education-Security" nexus may seem unthinkable. However, it exists and proves to be extremely significant.

Education, understood not only as a process, but as level of education - a set of knowledge, skills, competencies and upbringing, acquired and developed in the learning process, ceteris paribus, provides more opportunities for the realization of each member of society. It provides the opportunity to find a desired professional career both in terms of pay and in terms of professional and emotional satisfaction. Thus, in addition to being a source of self-esteem and confidence, education indirectly creates stability, life security, and is one of the factors for personal well-being. Ultimately, education is an important, of course, not the only, prerequisite for people to feel satisfaction with life and to be happy.

The concept of human security, first presented in Chapter Two, "New Dimensions of Human Security" of the 1994 UNDP Human Development Report, states that human security has two main components: freedom from fear and freedom from wants.

The same report defines seven "threat categories" or human security dimensions, such as:

- Economic security – It requires a basic income, usually from paid employment in the public and private sectors, as well as income from self-employment or at least from some publicly funded social network. In the early 1990s, only a quarter of the world's population could be called secure in this sense. The greatest threat to people's economic security is poverty.
- Food security – It means that all people at any time have physical and economic access to essential foods. This requires more than just distributing food. Very often there is food available, but the population does not have the financial resources to buy it.
- Health security – It refers to access to health care and health services for every person. The main threats are injuries and communicable diseases. This security is lowest in poor and rural areas. Women and especially children are among the most vulnerable groups.

- Environmental security – It requires healthy physical environment, ensuring the integrity of local ecosystems, the purity of air, water and soil, limiting the processes of deforestation, desertification, salinization, as well as combating natural disasters and minimizing the harmful effects associated with human activity.
- Personal security – It is one of the most important aspects of human security, directly related to ensuring the physical security of people. In both poor and rich countries, human life is increasingly threatened by unexpected and unpredictable forms of violence. Threats can take different forms: threats from one's own country (threat of torture), threats from other countries (interstate military conflicts), threats from other groups (ethnic tensions), threats from individuals or gangs to other individuals or gangs (organized crime, street violence), etc.
- Community security – It refers to the security of groups distinguished on different grounds - ethnic, cultural, religious, ideological-political, gender, etc. It finds expression in the observance of their group rights, sometimes in their physical survival.
- Political security – It means that people live in a society, in which their rights are respected and there are no threats of military dictatorship, arbitrariness and abuse, political repression, or state terror.[4]

Human security is a complex phenomenon and its dimensions are interrelated. However, it is clear that economic security, food security and, to a large extent, health security belong to the “free of wants” component. Their guarantee requires people to have resources, mainly financial ones. The main legitimate source of such finances is employment from the exercise of a profession, which also requires the appropriate education.

The relationship (causal relationship) between education and the three dimensions of human security – economic, food and health security – might also be seen in the Sustainable Development Goals (SDGs). The first three of them – SDGs 1. No poverty, SDGs 2. Zero hunger and SDGs 3. Good-health and Well being, are followed by the goal related to education – SDGs 4. Quality education. SDGs 4 seeks to ensure access for all to quality education through all stages of life, as well as to increase the number of young people and adults who have the relevant skills for employment, decent jobs, and entrepreneurship.[3]

The remaining dimensions of human security, which may be attributed to the other main component "free of fear", may also, although more indirectly, depend on education. Education is not just a process of mechanical accumulation of knowledge and development of skills. The accumulated knowledge, ceteris paribus, combined with its understanding, leads to the intellectual development of the personality. Education also implies upbringing, which, if based on basic universal human values and inherent virtues, should contribute to the moral elevation of each person. Thus, people would be more inclined to accept differences, respect others, seek to avoid violence, and not resolve their disputes through conflict. Thus, levels of individual, group, and political security within the concept of human security would increase.

The level of education as a characteristic of people, who are members of a society, is decisive not only for their personal security, as stated above, but also turns out to be a factor of national security, understood in the broadest sense.

While the size of a country's population is an essential quantitative characteristic, the level of its education is a key qualitative characteristic. A highly educated population in a country, combined with other resources, may contribute much more to its economic and social development, as well as to its overall well-being, than if it is poorly educated and illiterate.

Although it is difficult to determine which countries are currently the most educated, a number of organizations process statistical information related to the education of the population around the world. For example, according to a ranking based on data on the percentage of each country's residents between the ages of 25 and 64 who have completed some form of tertiary (higher) education, Canada ranks first. The share of the population with

tertiary education in this country is 65%. Canada is followed by Ireland - 58%, Japan - 57%, South Korea - 56%, Luxembourg - 54% and Australia - 53%. Bulgaria is in 30th place, having 34%. Germany has the same share. In the ranking of the population with secondary education, the situation is as follows: Czech Republic – 67%, Slovakia – 65%, Croatia – 60%, Hungary – 57%, Romania – 56%, Poland – 55%, Slovenia – 54%, Bulgaria – 53%, Germany – 50%, etc.[8]

3. Education Quality Dimensions

When discussing the "Security-Education" nexus, one must first consider the educational level of both the individual and the society within a country. This educational level is the result of the quality of education, understood as a process taking place within the educational system.

The quality indicators of the educational process, which might also be presented as quality factors, may generally be: teachers / lecturers, especially their professionalism, educational documentation, including curricula and syllabi, teaching methods, the organization of the educational process, which is directly related to its management, the material base and information resources, including the provision of textbooks and other literature used within this process. The words „teachers“ and “lecturers” may be used as synonymous, but to be precise they have different meaning. The teacher is a person, who teach especially at school. The lecturer is a person, who gives lectures, especially at university. That is way both are used.

In combination, all these things are an essential condition for the educational process to proceed at a high level. In addition, two more conditions must be added, which are no less important in order to talk about quality education - the characteristics of the learners, who are a key participant in the learning process, and the context, including national or international, in which this process takes place.

The problem of the quality of education is a subject of interest for many researchers and engages the attention of various international organizations, including UNICEF, OECD, etc.

In his monograph, researcher Robin Alexander analyses the topic of educational quality and its definition in great depth. For example, at the very beginning he points out five indicators of educational quality published in the OECD Report “Education at a Glance” from 2000 [1]. These five indicators are presented in a separate chapter and include: Context of education, Financial and human resources invested in education, Access to education, participation and progression. The learning environment and organisation of schools, Individual, social and market outcomes of education, Student achievement [2].

At the meeting of the International Working Group on Education, held in Florence, the UNICEF declared that the education quality includes:

„ Learners who are healthy, well-nourished and ready to participate and learn, and supported in learning by their families and communities;

„ Environments that are healthy, safe, protective and gender-sensitive, and provide adequate resources and facilities;

„ Content that is reflected in relevant curricula and materials for the acquisition of basic skills, especially in the areas of literacy, numeracy and skills for life, and knowledge in such areas as gender, health, nutrition, HIV/AIDS prevention and peace;

„ Processes through which trained teachers use child-centred teaching approaches in well-managed classrooms and schools and skilful assessment to facilitate learning and reduce disparities;

„ Outcomes that encompass knowledge, skills and attitudes, and are linked to national goals for education and positive participation in society [6].

The UNESCO Quality Education framework presents the main elements of education systems and demonstrates how they relate to each other, providing a tool for monitoring quality and analysing policy choices. The central part of this framework is teaching and learning processes, which are influenced by the following dimensions:

• Learner characteristics: characteristics of the learner and his/her household that may determine the effectiveness of the educational process;

• Enabling inputs: material and human resources, including management, that directly influence teaching and learning processes;

• Teaching and learning: the process of teaching and learning, including essential pedagogical factors;

• Outcomes: results of the learning process on the individual and society;

• Context: social, cultural, economic and political dimensions of the local context. This aspect cuts across the whole process from characteristics to outcomes [7].

4. Challenges to Education Quality

The challenges regarding quality of education should be related to the main among the above-mentioned indicators, such as teachers/lecturers, learners, additional conditions related to the implementation of the educational process and the existing context.

Teachers/Lecturers

The main duty of teachers/lecturers is to teach knowledge to students in certain disciplines and to contribute these learners to develop specific skills and competencies. These should be people, who not only possess knowledge and experience in the relevant discipline, but they must also be motivated to work dedicatedly, responsibly, honestly and with integrity. In other words, teachers/lecturers must demonstrate professionalism, understand primarily as: first, applying specific knowledge, possessing skills and competencies, and second, demonstrating certain behaviour based on ethical principles valid for those employed in a particular professional field and practicing a specific profession.

The professionalism of teachers/lecturers depends both on the state's will to have well-prepared, motivated and honest teaching staff, which it purposefully seeks to develop through the implementation of a relevant policy, and on the correct career orientation of the individual, on his desire and potential for realization in the field of education. Career orientation may also be part of the state's policy related to the development of a quality teaching staff at the very initial stage of the professional development of each teacher/lecturer.

Proper career guidance is important, but it is not the only condition for a country to have good teachers/lecturers. The policy in this regard should include activities and measures related to the motivation and development of teaching staff.

A large number of teachers/lecturers begin their teaching career, which can be combined with research, feeling a calling. These are people, who are mostly driven by an understanding to be useful to society. Therefore, the understanding of the importance of their work would lead to an increase in the prestige of their profession, which in turn would be an important non-material incentive.

Certainly, motivation should be strengthened through financial incentives, including those based on the results obtained in the work process. Teachers'/lecturers' salaries should correspond to the importance of the profession.

An important element of the policy related to the development of the teaching staff is the creation of conditions for improving its qualifications. This may be done by including

teachers/lecturers in various forms of postgraduate training/continuing education. This training would have an additional motivating effect, because teachers are people who understand the power of knowledge. It will be better if this upgrading training is accompanied by some additional material incentives.

The approach of the state and the corresponding policy is a general framework for people with different values, attitudes and characters, whose behaviour is always specific, unique. Therefore, despite the existence of good policies and measures for developing and maintaining highly qualified and motivated teachers/lecturers, the positive effect of them will depend significantly on personal characteristics and other subjective features - worldview, responsibility, maturity, combined with life experience, etc.

Learners

Learners are expected to demonstrate a responsible attitude and diligence towards their duties. They must attend classes regularly, as well as not being systematically late for them, thereby interrupting the teaching process. The students must also maintain order in the classrooms while classes are being held. They must complete the tasks assigned to them on time and of highest quality. The learners should independently develop their assignments assigned by the teachers/lecturers, respecting copyright and citation rules. They must not cheat during the tests and exams, using various electronic devices. They must comply with the rules of internal order in educational institutions, such as not demonstrating aggression, not smoking, etc.

The great challenge for learners, as well as for teachers/lecturers, is related to the uniqueness of each person. The learners are not machines that follow a certain program or commands. They are living, thinking and feeling beings going through different stages of their individual development. The learners may show sympathy or antipathy towards teachers/lecturers or other students, towards academic subjects, the way the learning process is conducted, etc.

The profile of learners is also complemented by the social environment they grow up and live in. It also includes the upbringing they receive in their families, which shapes their attitude towards education, the educational process and other participants in it.

All this makes the teachers'/learners' work extremely difficult. This difficulty lies in the fact that they must find an approach to arouse interest in as many students as possible, taking into account their age, knowledge, interests, emotional state. They must make them pay attention to the issues in the relevant discipline, to motivate them to seek additional information, to be active and to demonstrate high results.

Additional conditions related to the educational process (financial and material resources, educational documentation, etc.)

The additional conditions contribute to improving the quality of the educational process, but they are not sufficient for this by themselves. Their influence should be considered in combination and inseparably with the other factors.

Among these additional conditions, educational documentation, including curricula and syllabi, is extremely important because it sets the framework for education. It must correspond with the continuous development of theory and practice in the relevant scientific field and the requirements for knowledge and skills that every enlightened person must possess in connection with this development. This is why curricula and syllabi are subject to periodic updating, both in terms of their content and teaching methods.

Better facilities, ceteris paribus, have a motivating effect on both students and teachers/lecturers. In addition, they greatly facilitate the educational process.

Context (family, community, national, international)

The *international context* is mainly about the view, which is being imposed at a global level on the importance of education for the improvement and well-being of people, living on the planet. This includes, first, its recognition as an undeniable value and universal human right.

According to Article 26 of the Universal Declaration of Human Rights 1. Everyone has the right to education. Education shall be directed to the full development of the human personality and to the strengthening of respect for human rights and fundamental freedoms. It shall promote understanding, tolerance and friendship among all nations, racial or religious groups, and shall further the activities of the United Nations for the maintenance of peace [5].

One of the SDGs, mentioned above, is not just education for the people, living on the planet, but quality education for all.

The *national context* is primarily about the understanding of the importance of education for the security and development of the individual (each citizen), society and the state and, accordingly, its prioritization. This vision of education also determines the policy, which the state pursues in this sphere, consistent with its resource capabilities. It implies setting clear and realistic goals, defining responsibilities and allocating resources, implementing specific activities and exercising control.

The national context is not only about the implementation of a far-sighted policy in the field of education. This context is much broader and may include the overall environment and the processes, which take place in it - historical, psychological, social, cultural, technological, economic, etc., often influenced by trends observed at the supranational level. For example, a country may have developed a very good education policy that takes into account the importance of the main factors for high-quality education, but this policy may not be implemented in the way it was conceived and may not lead to the desired results. The reasons may be some of those mentioned above.

The *community and family context* may play a key role. Regardless of the national context, which may be generally unfavourable, the community and family environment may be a powerful stimulus for acquiring knowledge and, consequently, higher education. This same environment could also be an obstacle/“brake“ to development and education under generally favourable socio-economic conditions.

5. Conclusions

In conclusion, it may be emphasized once again that the education in general, and specifically the educational process and the educational system within which it takes place, are strongly related to security. The knowledge that is taught and acquired within the educational process is a valuable resource. It is a power that may belong to both the individual and the state. This power provides the opportunity to achieve the goals, which every the individual or a society sets for themselves.

There are various challenges facing education that are interconnected. One of them is the scarcity of resources and how to manage them in a way that maximizes the achievement of educational policy goals at the highest possible quality. Another challenge relates to the training and motivation of teachers/lecturers. Teaching staff have a key role within the educational process. Every teacher or lecturer faces thousands of learners from different generations during their career. He/she does not only teach knowledge. His/her most important task is to motivate learners to be active in the educational process and to realize the importance of education for their better future and for the future of the country.

Acknowledgement

The scientific publication is in fulfillment of the objectives of the National Scientific Program "Security and Defense", financed by the Ministry of Education and Science of the Republic of Bulgaria in implementation of the Decision of the Council of Ministers No. 731 of October 21, 2021 and according to Agreement No. D01-74 /19.05.2022.

References

1. Alexander R. (2017). Education For All. The Quality Imperative and the Problem of Pedagogy. CREATE PATHWAYS TO ACCESS Research Monograph No 20 [Online]. Available form: http://www.create-rpc.org/pdf_documents/PTA20.pdf
2. Education at a Glance - OECD Indicators 2000. Available form: https://www.oecd.org/en/publications/education-at-a-glance-2000_eag-2000-en.html
3. Eurostat. Quality education. Available form: https://ec.europa.eu/eurostat/web/sdi/database/quality-education
4. Ul Haq, M., Inge Kaul and contributors. (1994). Chapter Two. New Dimensions of Human Security, Human Development Report [Online]. Available form: https://digitallibrary.un.org/record/240220?v=pdf
5. UN (1948). Universal Declaration of Human Rights. Available form: https://www.un.org/en/about-us/universal-declaration-of-human-rights
6. UNICEF (2000). Defining Quality in Education. Working Paper Series Education Section. Available form: https://www.right-to-education.org/sites/right-to-education.org/files/resource-attachments/UNICEF_Defining_Quality_Education_2000.PDF
7. UNICEF (2017). Development of the National Quality Education Conceptual Framework and Design of its Monitoring System, Available form: https://www.unicef.org/uzbekistan/media/3661/file/QEF_Conceptual_Framework_eng.pdf
8. World Population Review. Most Educated Countries 2026. Available form: https://worldpopulationreview.com/country-rankings/most-educated-countries

AI for Detecting and Combating Deepfakes and Disinformation

Veronica Hîncu

Academy of Economic Studies of Moldova, Chisinau, Moldova
soltanici.veronica@ase.md

Abstract. Advances in generative technology now allow the creation of deepfakes - videos, images, and audio that appear authentic but are completely fabricated. While these tools have creative uses in entertainment and media, they also pose serious risks. Deepfakes can spread false information, support fraud, and undermine trust in the digital content people rely on every day. Interestingly, the same technology that enables deepfakes can also help detect them. Modern detection systems identify subtle cues that are hard for humans to notice, such as slight irregularities in facial expressions, inconsistencies in audio, or unusual patterns in video frames. More advanced approaches combine multiple signals—including visual, audio, and metadata analysis—while digital watermarking and provenance checks help confirm when content is authentic. Even so, detection is not perfect. New types of deepfakes can bypass existing systems, attackers can deliberately design content to fool detectors, and sometimes genuine material is mistakenly flagged as fake.

Because of these challenges, addressing deepfakes requires more than just technology. Clear laws, responsible online platforms, and stronger digital literacy are essential. By combining innovation with ethical responsibility, it is possible to protect information integrity and maintain trust in the digital world.

Keywords. Deepfake detection, generative AI, disinformation, cybersecurity, information integrity, ethical AI.

1. Introduction

In the quiet libraries of Chisinau, just a decade ago, the historical archives felt like sanctuaries of absolute truth. To the Moldovan citizen, a photograph in a newspaper or a snippet of film on the evening news was not just a representation of reality—it was reality itself. The core philosophy of public life was simple: *"seeing is believing."* This unspoken contract between media and the public formed the bedrock of shared understanding. When an official spoke, when a document was displayed, it was accepted as authentic because the alternative—that it was manufactured from nothing—belonged only in the realm of science fiction.

That contract has now been broken. We have crossed the threshold into the era of synthetic media, a paradigm shift so profound that it undermines the very concept of objective reality. In this new world, seeing can no longer be believing.

This disruption was accelerated by an explosion in generative artificial intelligence. The technical barriers that once required specialized Hollywood studios have vanished, replaced by open-source algorithms and accessible apps [2]. This "democratization of deception" was driven by two key technologies: **Generative Adversarial Networks** (GANs) and **Diffusion Models**. In a GAN, one AI learns to create a fake image while a second AI attempts to detect it; this constant "adversarial" race pushes the quality to hyper-realistic levels.

Diffusion models refined this further, learning to generate immaculate, high-resolution visuals from simple text prompts. Today, anyone with a smartphone can create a plausible fabrication of a world leader, a neighbor, or a colleague saying things they never said and performing actions they never did.

The consequences of this shift are not theoretical; they are already reshaping the fates of nations. In the Republic of Moldova, during critical electoral periods, the information space was deliberately flooded with synthetic disinformation [8]. Citizens were targeted not merely with false text articles, but with manufactured "evidence." Advanced voice-cloning applications were used to create fake audio recordings of pro-European politicians making unpopular statements, optimized specifically for messaging apps to sow confusion among older demographics. Elsewhere, deepfake videos appeared on social media, using GAN-generated faces to alter official debates or depict fictional crises. The technology was weaponized to inflame existing social tensions, discredit democratic institutions, and push nostalgic, pro-Russian narratives by exploiting familiar faces.

While deepfakes pose this systemic risk to information integrity, the same intelligence used to create them can also be harnessed to defend against them [9]. Modern detection systems identify inconsistencies invisible to the human eye, such as unnatural blood flow patterns on a synthetic face or irregularities in frame sequences [10]. However, detection alone is not a "silver bullet." It is an endless arms race, where new deepfake generators are constantly trained to bypass the latest detectors. Furthermore, the risk of false positives—accidentally flagging genuine footage as fake—can damage trust just as severely.

Ultimately, technological innovation must be matched by ethical and structural responsibility. A multi-layered approach—combining AI-driven detection, standardizing content provenance (tracking the origin of media from camera to consumer), and robust legislative policy—is the only viable defense. By restoring verification to the heart of the digital ecosystem, we can protect the integrity of our information and maintain a world where truth still has a foundation.

2. The Mechanics of Deception

If the Introduction established that we can no longer trust our eyes, Section 2 explores the "engine room" where that trust is systematically dismantled. The creation of a deepfake is not a simple filter or a clever edit; it is a high-speed digital duel, a constant evolution occurring at the speed of light [5].

2.1. The Internal Duel: The Forger and the Detective

Imagine a master forger and a seasoned detective locked in a room together. The forger draws a banknote; the detective immediately points out a flawed watermark. The forger tries again, correcting the error, only for the detective to find a microscopic bleed in the ink. This cycle repeats thousands of times, with the forger growing more skilled and the detective more observant, until finally, the detective can no longer tell the difference between the creation and the original.

In the world of Artificial Intelligence, this is the **Generative Adversarial Network** (GAN). Within this system, two neural networks are locked in a relentless race. The **Generator** (the forger) attempts to create synthetic data from scratch, while the **Discriminator** (the detective) evaluates the work against real-world data. Every time the detective catches a flaw, the generator evolves to bypass it. This is an automated evolution of deception, allowing AI to move beyond blurry, uncanny shapes into a realm of hyper-realism that defies human perception.

2.2. The Expanding Toolkit: Beyond the Face-Swap

The deception has matured, moving far beyond the "uncanny valley" of simple face-swaps. To understand the modern threat, we must look at the full spectrum of synthetic media—tools that are becoming increasingly democratized [1].

The new frontier is arguably Audio Cloning, perhaps the most dangerous weapon in the digital arsenal. With as little as thirty seconds of a target's voice, AI can now clone their vocal "DNA." This is not a mere recording; it is a digital instrument that can be played in real-time, capable of saying anything with the target's exact inflection, breath, and emotion. Alongside this, Full-Body Synthesis allows attackers to map the movements of a professional actor onto the body of a political leader, creating a "digital puppet" that moves with lifelike fluidity.

Yet, the most pervasive threat often comes from the "low-tech" end of the spectrum. "Cheapfakes"—videos manually slowed down to imply intoxication or cropped to strip away context—remain devastatingly effective. They require no supercomputers to create, only a malicious intent and a social media platform to spread them before the truth can catch up.

2.3. The Dark Incentives: The Weaponization of Reality

While the technology itself is neutral, the intent behind it is increasingly predatory. The "Mechanics of Deception" are currently fueled by three primary dark incentives:

1. **The Political IED:** As witnessed during recent electoral tensions in Moldova, deepfakes act as digital landmines. They are often timed as an "October Surprise"—dropped just hours before an election, leaving no window for the victim to debunk the footage before the ballots are cast. In a fragile information ecosystem, a well-timed fake can alter the course of a nation.

2. **The Financial Heist:** Deception has moved from the streets into the boardroom. "CEO Fraud" has evolved into a sophisticated game where audio-cloned voices call subordinates to authorize urgent, multi-million dollar transfers. It is the ultimate social engineering tool, leveraging the sound of authority to bypass security protocols.

3. **The Erasure of Reputation:** At its most intimate and cruel level, deepfakes are weaponized for personal harassment. The creation of synthetic "revenge porn" or fabricated evidence is used to strip individuals—often women and public figures—of their dignity, power, and standing in society.

In this forge, the tools of creation are indistinguishable from the tools of destruction. The question is no longer *if* a fake can be made, but *how* we can ever trust the real again.

3. Digital Detectives: Hunting for the Ghost in the Machine

If the generator is a master of the mask, the detective must learn to look at the "soul"—or rather, those biological and mathematical signatures that a machine, for all its processing power, cannot yet perfectly mimic. To understand this battle, we must look at how technology is being used to unmask the digital imposters that have recently haunted our own doorsteps in the Republic of Moldova.

3.1. The Involuntary Witness: Reading the Biological Pulse

The most striking frontier in detection is the realization that a deepfake, no matter how flawless its skin, is often "hollow." Humans are biological machines with messy, rhythmic systems that breathe and pulse.

One of the most powerful tools in a detective's arsenal is **Remote Photoplethysmography** (rPPG). This technology allows a computer to "see" a person's pulse through a standard camera lens. Every time your heart beats, blood flushes into your face, causing a microscopic change in skin color—a change invisible to the human eye but loud and clear to a sensor. While advanced 2026 generative models have begun to simulate a pulse, they often fail the test of **Spatial Blood Flow**. In a real human, the forehead, cheeks, and neck flush in a coordinated, anatomical pattern. In a deepfake, the pulse often appears as a uniform "flicker" or is entirely disconnected from the person's breathing, revealing a digital entity that mimics life without actually possessing it.

3.2. The Moldova Case Study: Glitches in the Matrix

In Moldova, we don't have to look far for examples of these "digital IEDs." During the 2024 and 2025 electoral cycles, the information space was flooded with fabrications designed to erode public trust. We saw deepfakes of President Maia Sandu ranging from the absurd—alleging a ban on traditional berry-infused tea—to the malicious "celebrity scandal" stories engineered by the *Matryoshka* bot network.

However, these "ghosts" left traces. Detectors unmasked them by hunting for **Spatial Inconsistencies**. In the "berry tea" videos, for instance, the ear symmetry was slightly off, and the lighting on the face failed to match the ambient environment of the Presidential Palace. Advanced AI tools, such as those deployed by the WatchDog.md community [7], identified Temporal Jitters—tiny, unnatural "shivers" in the video frames. These jitters occur when a synthetic face struggles to stay "anchored" to a real actor's movements, creating a subtle visual vibration that screams "forgery" to a trained algorithm.

3.3. The Multimodal Interrogation: When Sight and Sound Collide

A deepfake might look perfect in a still frame, but the deception often collapses the moment it speaks. Modern detection acts as a digital cross-examiner through **Multimodal Analysis**. Forensic AI now analyzes the physics of speech, checking if the movement of the lips (visemes) matches the precise phonetic sounds (phonemes). In many of the fake audio clips targeting Moldovan politicians, the "vocal DNA" was cloned with terrifying accuracy, but the AI caught the **Spectral Signatures**—metallic, mechanical frequencies that do not exist in human vocal cords. Furthermore, the "Liveness" test tracks involuntary micro-expressions of the eyes, known as saccades. While a human eye is constantly darting and reacting, AI-generated eyes often look "static" or move with a robotic regularity that triggers an immediate alert.

3.4. The Digital Birth Certificate: Proving the Truth

As this arms race intensifies, the strategy is shifting from *finding* fakes to *proving* what is real. This is the era of the **Digital Birth Certificate**.

By 2026, the **C2PA standard** [3] is becoming the gold standard for media integrity. Imagine a journalist in Chișinău filming a government briefing; the moment the "record" button is pressed, the camera cryptographically signs the file. This creates a "chain of custody." If that video is later intercepted and altered by a Kremlin-backed AI tool to change the official's words, the digital signature is broken. A small "CR" icon appears on the user's screen with a clear warning: *Metadata Inconsistent—Content May Be Altered*. In this new landscape, the "Ghost in the Machine" is being hunted by the very technology that gave it life, ensuring that even in an age of synthetic media, the truth still has a heartbeat.

4. The Human Firewall: Law, Literacy, and the Social Vaccine

While algorithms can build the walls of our digital fortress, the strength of the gate depends entirely on the person holding the key. As we navigate the complexities of 2026, we have arrived at a sobering realization: technology is a powerful shield, but it is not a cure. The final line of defense against the "digital ghosts" of disinformation is not found in a line of code or a complex algorithm, but in the collective resilience of a nation's people and its institutions. In the Republic of Moldova, this **"Human Firewall"** is no longer a theoretical concept—it is a living defense being tested and tempered by the daily realities of the digital age.

4.1. The Legislative Shield: Closing the No-Man's Land

For years, deepfakes existed in a legal wilderness, a "no-man's land" where attackers could manufacture chaos with total impunity. However, between mid-2025 and 2026, the legislative landscape in Moldova shifted from reactive confusion to decisive accountability.

At the heart of this shift is the **Center for Strategic Communication** (STRATCOM). No longer content with merely debunking lies after they've gone viral, the Center has evolved into a proactive early-warning system. Like a digital radar, it identifies the "Matryoshka" bot networks as they warm up their servers, neutralizing deepfake campaigns before they can reach a fever pitch in the public consciousness.

Furthermore, as Moldova marches toward EU membership, it has harmonized its laws with the **EU AI Act** [4]. By June 2026, the rules of the game are clear: all synthetic content must carry a label. This isn't just a suggestion; it is a legal mandate for transparency. Whether it is an AI-generated ad promising "guaranteed returns" or a satirical political parody, the viewer must be told they are engaging with a machine's imagination.

4.2. Digital Literacy: Training the "Internal Detector"

If law is the shield, then education is the training ground. In Chișinău and across the rural reaches of the country, a new kind of "civil defense" has emerged. While technology moves at the speed of light, education moves at the speed of conversation—one household at a time.

A **"Wait and Verify"** culture is taking root. Organizations like *StopFals* and government spokespeople have transformed from news providers into forensic instructors. Citizens are learning to spot the "tells" of a machine: the unnatural jitter in a video of a Prime Minister or the rhythmic, robotic pause in a cloned voice promising a 1,700 MDL pension bonus.

Crucially, the country is building **generational bridges**. While younger Moldovans navigate the web as digital natives, their parents and grandparents remain the primary targets for "vishing"—voice-cloned phone scams. New community programs are replacing "heuristic trust"—the tendency to believe a familiar voice—with a healthy, critical skepticism. The message is simple: if the news is shocking, verify it through the official channels of the state before you believe your ears.

4.3. The Liar's Dividend and the Future of Trust

Yet, as our defenses grow, a shadow remains: the **Liar's Dividend**. This is perhaps the most poisonous byproduct of the deepfake era. As the public becomes more aware of AI's capabilities, corrupt actors have begun to use that very awareness as a shield. When caught in a real scandal or a genuine recording of misconduct, they now simply shrug and claim, *"It's*

an AI fake." They exploit the existence of deepfakes to escape the consequences of the truth. This is why the "digital birth certificate"—content provenance—has become so vital. It is no longer just about catching the lies; it is about providing an immutable record that protects the integrity of the real.

4.4. The Human Firewall: Building Your Personal Toolkit

While the "Digital Detectives" hunt for ghosts in the machine at a systemic level, the front line of this war often passes through our own pockets—specifically, our smartphones. In 2026, being a responsible citizen means evolving from a passive consumer into an active participant in our own defense. We must develop a "social vaccine" composed of low-tech wisdom and high-tech habits.

In an era where a mere three seconds of audio can be used to clone a loved one's vocal "DNA," the most sophisticated defense is surprisingly human. Families across Moldova are now adopting the "Safe Word" strategy. Imagine receiving a frantic call from a grandchild claiming they are in trouble and need an urgent transfer of funds. Before panic sets in, you ask for the family password—a random, nonsensical phrase like "Lego Teapot" or "Green Walnuts." In the world of 2026, if there is no secret word, there is no trust. This simple, analog lock is the one thing a digital forger cannot pick.

When engaging in video calls that feel "uncanny," we can use the limitations of the machine against itself. Even the most advanced generative models in 2026 often struggle with occlusions and complex 3D physics. If you suspect you are talking to a digital puppet, ask the person to turn their head to a full profile or wave their hand rapidly in front of their face. A deepfake "mask" will often glitch, flicker, or "melt" as the software fails to keep the synthetic face anchored to the physical reality of the movements. These visual stutters are the cracks in the digital illusion.

Deception thrives on emotional "high-resolution." Deepfakes are designed to bypass our logic and hit our lizard brains—fear, anger, or extreme excitement. The narrative defense begins with an internal audit: *Does this feel too perfect?* If a politician is making a radical statement in a high-end studio setting while they are supposed to be at a chaotic protest, the context doesn't fit the content. Our intuition—that subtle sense that something is "off"—is often our first and most reliable alert system.

Finally, we must move from verifying *what* we see to *where* it came from. Before hitting "share" on a shocking clip from an anonymous Telegram channel, look for the "Digital Birth Certificate." By 2026, major platforms have integrated the Content Credentials (CR) icon. A simple click allows you to see if the media was cryptographically signed by a reputable news agency like *Moldpres* or if it has been flagged as "Altered" by independent fact-checkers at *StopFals.md*. In the age of synthetic media, an unverified source should be treated as radioactive—potentially dangerous and best handled with extreme caution.

5. Conclusion

The Republic of Moldova is currently adjusting its national security protocols to address the increase in synthetically generated media. This process has moved beyond the deployment of software and now includes a reorganization of institutional and social responses. In 2026, the data indicates that while technical tools serve as a primary layer of defense, they operate as one part of a larger system. The security of the information environment depends on the active coordination between state agencies, technology developers, and the public.

The interaction between generation and detection remains a technical cycle. As researchers deploy new detection methods, such as rPPG analysis to identify blood flow patterns or C2PA protocols for tracking content origin, generative models are updated to minimize these detectable signatures. Because this cycle is continuous, the national strategy has shifted from attempting to eliminate deepfakes to managing their impact through an integrated framework.

This framework currently operates through three primary actions:

1. **Automated Monitoring:** The use of algorithms to scan digital platforms for content that lacks verifiable metadata or displays structural inconsistencies.
2. **Regulatory Enforcement:** The application of the EU AI Act and national laws to require that all synthetic media carries a digital label, providing a legal basis for removing unlabelled content.
3. **Direct Education:** The training of citizens to apply a verification sequence—checking sources and testing for glitches—before accepting digital information as factual.

The challenge of deepfakes is now a permanent factor in the digital environment. The goal for Moldova is to maintain a functioning information ecosystem where verification is the standard response to new data. Success in this area is determined by the speed at which discrepancies are identified and the consistency with which the population applies critical analysis. Ultimately, the stability of the information space relies on the sustained participation of an informed public.

References

1. Baddi M., A. A. (2023). The Art of Cyber Defense: From Risk Assessment to Threat Intelligence. CRC Press,.
2. Btissam Acim, M. B. (2025, October 2). A Decade of Deepfake Research in the Generative AI Era, 2014–2024: A Bibliometric Analysis. Retrieved December 10, 2025, from https://doi.org/10.3390/publications13040050
3. C2PA. (2025, December 22). Technical Specifications for Digital Content Provenance v2.1. Documentation on cryptographic signing for media integrity. Retrieved from Coalition for Content Provenance and Authenticity: https://c2pa.org/
4. Commission European. (n.d.). The EU AI Act: Transparency Obligations for Generative AI Systems. Retrieved February 25, 2026, from https://digital-strategy.ec.europa.eu/en/policies/code-practice-ai-generated-content
5. Gehl R. W., S. T. (2022). Social Engineering: How Crowdmasters, Phreaks, Hackers, and Trolls Created a New Form of Manipulative Communication. MIT Press,.
6. Raport anual privind amenințările hibride și integritatea spațiului informațional. Monitorul Oficial al Republicii Moldova. (2026). Retrieved from CENTRUL PENTRU COMUNICARE STRATEGICĂ ȘI CONTRACARARE A DEZINFORMĂRII: https://stratcom.md/
7. Rusu, A. (August 2025). REȚEAUA RUSĂ DE DEZINFORMARE ÎN MOLDOVA: ANATOMIA UNEI OPERAȚIUNI PE REȚELELE DE SOCIALIZARE. Chișinău: WatchDog.md.
8. Sârbu, A. (2025, September 24). Elections Under Siege: Moldova's Battle for Truth. Retrieved December 15, 2025, from https://iwpr.net/global-voices/elections-under-siege-moldovas-battle-truth
9. Schick, N. (2020). Deepfakes: the coming infocalypse. New York: Twelve.
10. Sumon, M. S. (2026, February 21). Attacking Deepfake Detection Systems: How Adversaries Evade AI-Based Fake Media Detectors. Retrieved February 23, 2026, from https://medium.com/: https://medium.com/@shafiqulsumon007/attacking-deepfake-detection-systems-how-adversaries-evade-ai-based-fake-media-detectors-20317fe1537c

Some Ways to Develop a Culture of Cyber Security among Teaching Staff

Liubimir Chiriac, Violeta Bogdanova

"Ion Creangă" State Pedagogical University, Chisinau, Moldova
llchiriac@gmail.com, bogdanovaleta@gmail.com

Abstract. This paper examines the current needs of teaching staff to improve their cyber security skills. The need to master informal cyber security tools alongside formal ones is becoming increasingly evident. Significant advances in information technology and the increasing number of cyber threats directly impact both educational organizations and their participants. Ways to foster a culture of cyber security among teaching staff to reduce the risk of financial and reputational losses are proposed.

Keywords. Cybersecurity, information security culture of teaching staff, reputational losses of teaching staff, cyber threats in education.

1. Introduction

At the current stage of development of the information society, issues related to the protection of computer information are of concern to both individuals and society as a whole. At the national level, the risks of information leakage can lead to serious consequences for national security, instability, and reputational damage.

Examples of information security breaches at the state level include:

- defense;
- data on strategic reserves of raw materials, precious metals, and diamonds;
- information on the protection of critical information infrastructure;
- compromise of the information and telecommunications infrastructure of government agencies and institutions;
- leakage of citizens' personal data;
- national security, etc.

For commercial organizations, the implementation of information threats most often leads to financial losses, disruption of the organization's core activities, and reputational damage.

Non-profit organizations also process confidential information, such as:

– Adoption confidentiality;
– Medical confidentiality;
– Court and investigation confidentiality;
– Tax confidentiality;
– Banking confidentiality;
– Attorney-client confidentiality;
– Notarial confidentiality;
– Employee personal data;
– Trade secrets, etc.

Today, matters related to official secrets are the responsibility of specific organizations. Legislation does not define the relationship between state secrets and other types of secrets,

such as commercial secrets and professional secrets. The issue of legal protection of official secrets has attracted considerable attention from researchers in the fields of law, management, and information security. [1].

For individuals, the consequences of negative cyber attacks lead to financial and reputational losses. While it was previously believed that school-age children and older adults were the most vulnerable, today, cybercrime victims can include people of all ages and professions, including cybersecurity specialists.

Personal data typically includes:

– Last name, first name;
– Telephone number, email address;
– Date and place of birth;
– Residence address (permanent or temporary registration, actual residence);
– Passport information;
– Driver's license information;
– Information about employment, education, marital status, and property;
– Bank details;
– IP address, cookies, user geolocation, etc.

Data breach incidents are constantly covered in the media. Organizational solutions for data access control and the use of software and hardware for information security help reduce the likelihood of information threats. However, the human factor negatively impacts information security. Information protection and information security are lagging behind the pace of digitalization. Moreover, many companies, at the current stage of economic and social development, are paying insufficient attention to this. [2].

The education sector is no exception. Due to certain specific characteristics, educational institutions often become training grounds for attackers.

Numerous educational institutions around the world are being hacked, resulting in reputational and financial losses. [3-6].

Cybersecurity issues are currently addressed quite effectively at the administrative level. Educational institutions are creating IT departments dedicated to cybersecurity. Students in certain majors are taught the fundamentals of cybersecurity. Schools offer a module on online security. Faculty are among the most vulnerable when it comes to cybersecurity skills. This article discusses ways to improve faculty skills in the field of information security.

The purpose of this work is to study the role of the information culture of the teacher in ensuring the information security of the educational organization.

The main objectives of the work:

- Analysis of cyber risks faced by teaching staff in a modern educational institution;
- Determining the role of information culture in an educational institution's information security system;
- Developing proposals for developing teacher information culture to improve the effectiveness of the educational institution's information security system.

2. Information culture of the teacher

All participants in the educational process are exposed to cyber threats in an educational institution: students, teachers, and administration.

Due to certain financial and technical constraints, educational organizations often first actively digitalize all processes, followed by (much less intensively) the development of an information security system. Unlike commercial entities such as financial institutions and telecommunications companies, educational institutions lack the ability to implement expensive hardware and software security measures, nor to strictly restrict access to

information resources and systems. Organizational information security measures based on international standards (ISO/IEC 27001:2022, ISO/IEC 27002, ISO/IEC 27005:2022, and others) and best practices are often a more cost-effective solution.

Considering that a technical support service operates in the interests of the administration, and new modules are periodically introduced for students, and master classes on personal data protection are held, the question of teachers' competence in the field of information security remains open [7].

In our view, developing teachers' information security competencies requires more than just conducting training seminars and implementing educational institution information security regulations. It is essential to foster an information culture among teachers..

In the work of Choo, Chun Wei, et al (2008) information culture is considered from 4 directions:

– result-oriented culture,
– rule-following culture,
– relationship-based culture,
– risk-taking culture [8].

Artemieva V.V. Voronina L.V. Utyumova E.A. (2020) identify 4 components in information culture: «*Information culture of the teacher as a personal integrative quality that allows the teacher to interact in the information educational environment based on the use of both traditional and new information technologies and represents the result of the interaction of:*

– *value-assessing,*
– *cognitive-communicative,*
– *effective practical and*
– *reflexive-evaluative components*» [9].

Information culture is a broader concept than digital culture, combining work with digital technologies when working with information and traditional [10].

Despite the various approaches to defining information culture, researchers agree that developing an employee's information culture leads to the overall success of the organization. Therefore, developing an information culture in the field of information security is an important element of ensuring an information security system.

3. Methods for developing digital culture among teachers

In a modern educational institution, all primary and secondary business processes are being digitalized [11]. The teacher is most often involved in the teaching-learning-assessment process.

Developing a teacher's digital competence begins in the educational institution during professional training. Older teachers (65+) did not have the opportunity to study the fundamentals of information technology in college or university. Middle-aged teachers studied computer science and ICT, but did not study information security issues. Unfortunately, ICT training programs for pedagogical majors do not always address information security issues.

Another approach to fostering a sound attitude toward information security is through professional development courses and massive open online courses. Continuing education often focuses on training teachers to create a safe environment for students..

In our opinion, it is important to include legislative and organizational components in the system for developing an information culture in the field of information security.

The following legislative acts in the field of information security exist in the Republic of Moldova:

1) Law No. 48 of 16-03-2023 on cybersecurity (published: 28-04-2023 in the Official Gazette No. 151-153 art. 225) – «regulates the regulatory, organizational and cooperation framework in the field of cybersecurity, establishes the competence of public authorities and institutions in the field of cybersecurity, determines the general national framework for crisis management in the field of cybersecurity, establishes requirements, measures and mechanisms for the purpose of ensuring the security of networks and information systems, which are essential for the functioning of society, and the management of cyber incidents»;

2) Law No. 72 of 10-04-2025 on electronic communications – establishes the national regulatory framework for activities regarding electronic communications networks and services, associated facilities and associated services, as well as certain aspects of terminal equipment, by defining the attributions of the relevant bodies in the field of electronic communications, the rights and obligations of natural and legal persons in the process of creating, managing and using electronic communications networks;

3) Law no. LP148/2023 of 09.06.2023 on access to information of public interest – «regulates: a) the manner of exercising and defending the right of access to information of public interest; b) the obligations of information providers in ensuring access to information of public interest; c) legal liability for violating the provisions on access to information of public interest»;

4) Law No. 142 of 19-07-2018 on data exchange and interoperability – «regulates the relationships that arise in the process of data exchange between information systems owned by public authorities and institutions, autonomous administrative authorities, as well as other organizations indicated in paragraph (2), on the one hand, and, on the other hand, information systems owned by private legal entities and individuals, the latter to the extent that they manifest their intention to voluntarily participate in the data exchange»;

5) Law No. 467 of 21-11-2003 on informatization and state information resources (published on 01-01-2004 in the Official Gazette No. 6-12 art. 44) – « establishes the basic rules and conditions of activity in the field of creation and development of the national information infrastructure as a functioning environment of the information society of the Republic of Moldova, regulates the legal relations that arise in the process of creation, formation and use of automated state information resources, technologies, systems and information networks»;

6) Law No. 195 of 25-07-2024 on the protection of personal data (published: 23-08-2024 in the Official Gazette No. 367-369 art. 574) – «establishes the legal framework regarding: a) the protection of individuals with regard to the processing of personal data; b) the organization and functioning of the National Center for Personal Data Protection (hereinafter referred to as the Center); c) supervision of the implementation of the legislation on the protection of personal data; d) legal liability and sanctions applied for violation of the legislation on the protection of personal data» ;

7) Law No. 982 of 11-05-2000 on access to information (published: 28-07-2000 in the Official Gazette No. 88-90 art. 664) – «regulates: a) the relations between the information provider and the natural and/or legal person in the process of ensuring and exercising the constitutional right of access to information; b) the principles, conditions, ways and means of achieving access to official information held by information providers; c) aspects of the accessibility of personal information and its protection in resolving the access issue; d) the rights of information seekers, including personal information; e) the obligations of information providers in the process of

ensuring access to official information; f) the manner of defending the right of access to information».

Understanding that information security issues are not only an actual agenda, but also an area regulated by legal acts, will allow teachers to form a legal culture when working with information.

Acquaintance with standards in the field of information security will allow the teacher to understand that the information protection measures applied in the educational organization are recognized as a necessary limitation, allowing to ensure the integrity, confidentiality and availability of valuable information.

4. Conclusion

The unprecedented growth of information threats and cyber incidents affects not only financial institutions and government agencies, but also the educational environment. The risks, threats, and losses of an educational institution are similar to those of any other organization. Knowledge of legislation and international standards related to information security is essential for modern educators and educational institutions to protect their information assets. (Figure 1).

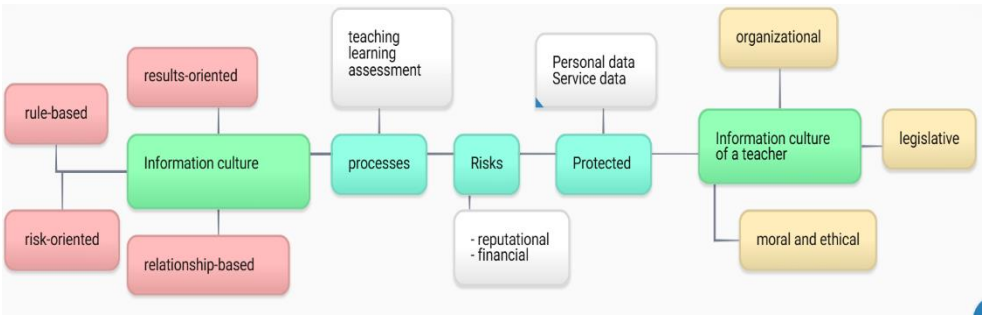


Figure 1. Conceptual framework for developing a teacher's information culture
* *created by the authors in the online service bubbl.us*

The work clearly shows the connection between the information culture of the teacher in the field of information protection and the improvement of the efficiency of the work of the educational organization from the position of reducing reputational and financial losses due to the realization of cyber incidents.

References

1. Karmanovsky, N.S., Savchenko-Novopavlovskaya, S.L. Organizational and legal foundations for the protection of official secrets at the enterprise St. Petersburg: ITMO University, 2018.
2. Quyet, Nguyen Xuan, and Truong Kim Phung. "Comparative analysis of information security policies at Big 4 Vietnamese logistics companies.“." International Journal of Multidisciplinary Reseach and Growth Evaluation 4.6 (2023): 683-690.
3. Official website of Washington Square News. https://nyunews.com/news/2025/03/22/nyu-website-hacked-data-leak
4. Official website of cyberinsider. https://cyberinsider.com/columbia-university-suffers-data-breach-affecting-870000-individuals
5. Official website of sttinfo. https://www.sttinfo.fi/tiedote/70198533/
6. Official website of bbc. https://www.bbc.com/news/uk-england-manchester-66005335

7. Bogdanova, V. Importanța dezvoltării unei culturi a securității cibernetice în rândul personalului didactic. Abordări inter/transdisciplinare în predarea ştiinţelor reale, (concept STEAM), Chişinău: CEP UPSC, 2025.
8. Choo, Chun Wei, et al. "Information culture and information use: An exploratory study of three organizations." Journal of the American society for information science and technology 59.5 (2008): 792-804.
9. Artemieva, V. V., L. V. Voronina, and E. A. Utyumova. "Development of information culture of future teachers in the conditions of digitalization of education." 2nd International Scientific and Practical Conference "Modern Management Trends and the Digital Economy: From Regional Development to Global Economic Growth"(MTDE 2020). Atlantis Press, 2020.
10. Curry, Adrienne, and Caroline Moore. "Assessing information culture—an exploratory model." International journal of information management 23.2 (2003): 91-110.
11. Chiriac L., Bogdanova V. Process Approach to Cyber Risk Management in Educational Organizations In: 14th International Conference on Application of Information and Communication Technology and Statistics in Economy and Education (ICAICTSEE – 2024), December 2-3RD,2024,UNWE, Sofia, Bulgaria, 2025.

Artificial Intelligence in the Theory and Practice to Consumer Behavior

Irina Balina

Academy of Economic Studies of Moldova, Chisinau, Moldova
555balina@gmail.com

Abstract. The article substantiates the relevance of the research problem and outlines the main research objectives. The current state of affairs and the specific features of 2025 are examined. Based on real-world examples, the study demonstrates how artificial intelligence is actively used in the production of consumer goods, as well as in the theory and practice of consumer behavior. A classification of the opportunities and challenges associated with artificial intelligence is presented, along with the principal international and regional regulatory acts governing AI. An analysis of the key directions of artificial intelligence regulation in the Republic of Moldova is conducted. Conclusions and recommendations are formulated.

Keywords. Artificial intelligence, consumer behavior, algorithms, entrepreneurial practice, Internet of Things.

1. Introduction

The relevance of the problem and the main research objectives can be formulated as follows:

- Consumer behavior analysis based on artificial intelligence methods enables obtaining profound insights into purchasing patterns, preferences, and trends. The outcome of such research is the development of a data-driven approach aimed at optimizing the effectiveness of digital marketing in terms of customer acquisition.
- Artificial intelligence algorithms create opportunities for analyzing customer data and behavioral patterns, identifying unique preferences, and accordingly adapting product recommendations or consumer-oriented communications. This level of personalization not only increases customer satisfaction but also improves conversion rates and customer loyalty.
- In this context, the study of the specific features of AI implementation in entrepreneurial practice appears highly relevant and in demand, both from scientific and applied perspectives.

Overall, despite the promising potential of artificial intelligence, today it has divided scholars worldwide into two opposing camps: some are convinced of its benefits, while others warn about its potential to destroy humanity.

2. Current State of Affairs. Specific Features of 2025

Artificial intelligence technologies and AI-based products exist in a wide variety of forms — from text editors and text-to-speech software to all devices connected through the Internet of Things (IoT). AI can be found in almost every industry where digital technologies are integrated into everyday life [1].

It is expected that by 2025 the number of artificial intelligence users worldwide will reach 378 million, while the AI market volume is projected to amount to USD 244 billion.

According to the World Economic Forum, legislative bodies around the world have adopted laws related to AI, reflecting concerns about the growing influence of artificial intelligence, particularly regarding algorithmic bias and data privacy. There are also concerns that AI may lead to the disappearance of certain professions in marketing, especially entry-level positions focused on the creation of basic content [2].

3. AI in the Theory and Practice of Consumer Behavior

Real-world examples demonstrate that AI is actively used in the production of consumer goods.

Theory: The consumer goods industry, which for a long time relied on established processes and brand loyalty, is undergoing radical transformation driven by artificial intelligence. While AI is often viewed merely as a tool for optimizing operations and supply chains, its true potential lies in its ability to fundamentally transform the processes of product development, promotion, and sales.

Practice: Examples of Application

Speak AI can be used to analyze customer interviews, identify sentiments and recurring themes, transcribe and evaluate focus group discussions, and provide actionable recommendations. The platform also generates comprehensive qualitative reports for strategic decision-making.

From a theoretical perspective, generative AI can synthesize realistic data and extract valuable insights from vast amounts of unstructured information, making it indispensable for targeted marketing, customer service improvement, and operational optimization.

Depending on the type of content required, various generative AI models may be employed, including Generative Adversarial Networks (GANs), Variational Autoencoders (VAEs), and transformer-based models.

In practice, the use of generative AI for content creation enables marketers to achieve more in less time: from generating social media posts and entertainment content to transforming images into trending TikTok memes, creating Instagram stickers, responding to requests through chatbots, supporting professional development and recruitment processes on LinkedIn, and brainstorming ideas for YouTube videos.

From the perspective of personalization, instead of relying solely on demographic data, companies can tailor experiences to the needs of each individual customer. For example, the cosmetics brand Sephora uses AI-powered chatbots to answer customer questions and provide personalized beauty recommendations. Healthcare companies employ generative AI to develop individualized treatment plans and improve patient care.

Voice assistants and smart devices are becoming increasingly widespread in practice. Behind the apparent simplicity of recognizing human speech and executing user commands lie natural language processing (NLP) and machine learning algorithms that enable AI to understand requests and context. These systems continuously analyze spoken language to improve recognition accuracy and response quality. AI training is based on real dialogues and adapts to accents, intonations, and speech patterns of different users. It is precisely this capacity for self-learning that makes voice assistants convenient and highly functional tools in consumers' everyday lives.

AI models such as ChatGPT-4o and Claude 3.5 Sonnet are capable of controlling third-party applications through voice interaction, not only executing commands but also maintaining more meaningful conversations, understanding complex requests, and even

demonstrating elements of empathy in their responses. This makes interaction with AI more natural and intuitive for users.

Today, AI applications are already capable of performing tasks such as:

- generating posts, visuals, and other marketing materials;
- reading news, work reports, and meeting notes aloud;
- creating deadline reminders;
- adding events to calendars, sending messages, making callback requests, etc.

The creation of such assistants can easily be implemented through Telegram bot development by specifying the required functionality in prompts.

Recommendation systems in online stores represent a practical demonstration of AI's role in e-commerce. This is one of the classical applications of artificial intelligence. AI is used to recommend the most relevant products by analyzing users' browsing and purchase histories, preferences, and the behavior of similar customers.

Such recommendation systems significantly improve user experience by helping consumers find relevant or interesting products among vast assortments. They also contribute to increased sales for retailers. Behind these systems are sophisticated machine learning algorithms that analyze large volumes of user behavior data. The use of AI in this context benefits both sellers and buyers.

The personalization of offers through AI assistants is another important direction in the use of artificial intelligence in e-commerce. Here, AI enables not only the adaptation of product recommendations, as is commonly assumed, but also the generation of website content, advertising banners, and email campaigns tailored to consumer needs.

For example, AI assistants from GPTunneL, based on ChatGPT, Gemini, or Claude, are frequently used to generate personalized content according to predefined prompts:

- selecting a model and configuring settings in just a few clicks;
- connecting to the vector database of a required marketplace;
- integrating APIs of necessary services, such as platforms for user data analytics;
- sending requests through the assistant's chat interface [3].

Taking into account both the advantages and disadvantages of AI technologies, it can be stated that they may prove indispensable in the following areas:

- **Error reduction.** With proper programming, errors can be minimized to nearly zero. A notable example is robotic surgical systems capable of performing complex procedures with exceptional precision.
- **Risk reduction for consumers.** AI-powered robots can perform hazardous tasks instead of humans, including work in harmful environments, critical situations, and disaster-response operations.
- **High operational efficiency.** AI systems do not require rest or lunch breaks, do not become ill, and are capable of continuous operation.
- **Business assistance.** Many companies and online stores employ AI technologies in the form of chatbots and virtual assistants.
- **Lack of emotional bias.** Artificial intelligence is devoid of emotions, functioning in a rational and practical manner. This characteristic may be effectively applied in personnel recruitment processes, where decisions can be based exclusively on applicants' professional qualifications and competencies.
- **Capability to perform routine tasks.** AI systems are highly effective in handling repetitive operations such as document verification, error detection, and mass email distribution. By delegating routine activities to AI, individuals can focus on business activities requiring creativity and strategic thinking [4].

Thus, such assistants can be configured for marketing, copywriting, design, product card creation, and many other tasks.

4. Opportunities and Challenges of Artificial Intelligence

At present, it can be stated that artificial intelligence plays a substantial role in both the theory and practice of consumer behavior. For example, while AI enhances the efficiency and productivity of marketing activities, it simultaneously exerts a transformative influence on individual professions such as copywriting and design, contributing not only to their development and improvement but also, in the long term, to the potential disappearance of certain occupations.

Today, society faces both significant opportunities and serious challenges that require solutions and the establishment of responsible and ethical approaches to the use of AI technologies.

In this context, the principal opportunities include:

1. Increased efficiency and productivity.
2. Expansion of innovation and creativity.
3. Automation and business growth.
4. Higher return on investment.
5. Stronger customer relationships.
6. Increased brand loyalty.

A particularly noteworthy development in the practical application of AI in consumer interaction is the emergence of so-called hyper-targeted campaigns.

In conditions of information overload, AI is regarded as a revolutionary tool for recommending products to consumers by delivering relevant information to the right audience at the right time.

Consequently, AI enables the reduction of informational “noise” and facilitates the promotion of hyper-targeted campaigns operating online on a 24/7 basis. Industries such as healthcare, education, and entertainment are capable of offering personalized product recommendations and dynamic email content that enhance customer engagement, as well as chatbots designed for individualized communication tailored to specific needs and preferences. By analyzing user preferences and behavioral patterns, AI algorithms recommend products and relevant content, thereby creating a seamless and personalized consumer experience.

Examples include Netflix and Amazon — leading streaming platforms operating independently yet technologically interconnected, as Netflix relies on Amazon Web Services (AWS) cloud infrastructure. Netflix is more strongly associated with original entertainment content, whereas Amazon Prime Video is frequently bundled with subscription services related to fast product delivery.

Netflix aggregates vast amounts of customer data, including search queries and viewing histories, and through AI-driven analytics generates recommendations tailored to individual preferences.

Similarly, Amazon employs AI-powered recommendation engines that analyze purchase histories, browsing behavior, and demographic information in order to provide personalized product suggestions likely to interest potential consumers [5].

At the same time, several substantial challenges are emerging:

1. The rapid pace of AI development and the risk of lagging behind technological progress.
2. Excessive dependence on automation.
3. Misuse and abuse of AI technologies, including the creation of fake content and copyright infringement.

4. Threats to privacy and security.
5. Insufficient resources and expertise for the full utilization of AI capabilities.

Particular attention should also be paid to another critical issue — the practical absence of comprehensive international legal regulation of artificial intelligence. At present, the regulatory framework remains in an active stage of formation, gradually evolving from ethical principles toward legally binding norms. By 2025, the dominant trend has become the establishment of risk-oriented legislative frameworks aimed at ensuring safety and the protection of human rights [6].

The principal international and regional regulatory instruments include:

1. **The EU AI Act (European Union Artificial Intelligence Act):**
    - Adopted in 2024, with phased implementation and full enforcement expected by 2026–2027.
    - It represents the world's first comprehensive legal framework regulating the development and use of AI systems.

The Act introduces the following classification of risks: prohibited systems (unacceptable risk), high-risk systems (e.g., healthcare and education requiring strict compliance procedures), and low-risk systems subject to minimal requirements.

2. **The Council of Europe Framework Convention on Artificial Intelligence** — the first legally binding international treaty aimed at safeguarding human rights, democracy, and the rule of law in the use of AI technologies.
3. **UNESCO Recommendations on the Ethics of Artificial Intelligence** — adopted by 193 countries, these recommendations emphasize human-centered approaches, safety, transparency, and data protection.
4. **OECD AI Initiatives** — containing AI principles that have served as the foundation for numerous national AI strategies [7].

The key principles of AI regulation include:

- **Safety and transparency:** AI systems must be auditable, predictable, and protected against unauthorized interference.
- **Accountability:** AI itself is not recognized as a legal subject; liability for damages rests with developers or users.
- **Non-discrimination:** The necessity of preventing algorithmic bias.
- **Privacy protection:** Compliance with data protection rights and standards comparable to the GDPR [8].

Legislative development increasingly follows the path of strict regulation of high-risk AI systems, particularly in the fields of biometrics, critical infrastructure, and personal data processing.

5. Main Directions of Artificial Intelligence Regulation in the Republic of Moldova

From a global perspective, the concept of AI development represents a major technological breakthrough. Artificial intelligence contributes to the integration of innovative technologies into socially and economically significant sectors of the Republic of Moldova.

From the consumer's perspective, AI creates opportunities for personal robotic assistants capable of performing household functions, such as controlling lighting and ventilation systems, scheduling tasks, reporting traffic conditions, or recommending nearby establishments based on individual preferences.

Facial recognition technologies may potentially replace traditional payment cards. The possibility of conducting payments in supermarkets through facial identification systems is currently being explored, which could eliminate the need for standing in queues.

Chatbots are becoming increasingly sophisticated and widespread. Already today, they are replacing human operators and consultants in online stores. From an economic standpoint, such systems significantly reduce labor costs, as they do not require salaries.

The development of "smart home" technologies is continuing, accompanied by an expansion of intelligent interconnected systems. Internet networks already connect hundreds of thousands of devices throughout Moldova, including not only mobile phones, tablets, and personal computers, but also household devices such as vacuum cleaners, televisions, air conditioners, and heating systems. The concept of the "Internet of Things" (IoT), where devices interact both with the Internet and with one another, is rapidly gaining popularity.

AI technologies are also actively used in translation services. Moldova entrepreneurs report successfully employing the HeyGen platform, which enables not only the translation of laws and articles but also video materials.

Innovations in AI are reshaping labor market dynamics. Two Moldovan entrepreneurs have developed an unconventional solution to one of the country's major challenges — labor shortages — through the use of holographic technologies and artificial intelligence. Their business operations are supported not by conventional employees but by virtual assistants. Such applications are also being implemented in healthcare services, while museums increasingly replace traditional guides with holograms of architects or historical figures associated with the exhibited sites [9].

AI applications in Moldova television media are also expanding. The news portal *Noi.md* became a pioneer in Moldova by experimenting with artificial intelligence in the production of studio-quality news videos. A virtual news presenter, Snejana Moralkina, generated through advanced AI technologies, delivered reports on national events in an ironic and engaging manner. It is important to recognize that the production of video content is traditionally a complex and expensive business process. AI technologies, however, make it possible for media resources of different scales — both large and small — to produce high-quality news and reporting videos. Developers at Mail.ru Group have suggested that "digital humans may become part of everyday life: virtual news presenters, virtual administrators, or even AI-generated movie stars" [10].

The use of virtual reality is likewise becoming increasingly commonplace. Although VR headsets are generally associated with entertainment, their potential applications extend far beyond this sphere, encompassing telecommunications, education, healthcare, tourism, and sports.

Table 1. Main Parameters of Artificial Intelligence Regulation in the Republic of Moldova by Areas of Application

Application Area	**Current Challenges**	**Importance of Regulation**
Healthcare	AI diagnostic errors	Mandatory algorithm verification and transparency
Education	AI-assisted plagiarism and dishonest use of AI in assignments	Clear usage standards and accountability
Finance	Manipulation in credit decision-making algorithms	Regulation of personal data protection and liability
Public Administration	Lack of transparency in decision-making	Mandatory reporting standards

Agriculture	Errors in automated systems	Safety and testing requirements
Industry	Risks to workers caused by automation	Safety standards and regulations (GOST)
Telecommunications	Collection and analysis of data without user consent	Legislative restrictions and sanctions
Transport	Autonomous control systems operating without supervision	Registration and licensing of projects
Marketing	Unfair targeting and exploitation of user profiles	Ethical standards and penalties
Cybersecurity	Vulnerabilities of AI systems	Strengthening protection and system monitoring

However, it should be recognized that alongside all the perespectives associated with the implementation of artificial intelligence, the Republic of Moldova also faces a number of significant disadvantages and challenges:

1. **Macroeconomic challenges**, particularly significant in the context of the existing budget deficit and growing public debt of the Republic of Moldova, including the necessity for substantial expenditures on software and hardware infrastructure, as well as the increase in unemployment and the gradual disappearance of certain professions.
2. **Psychological and social challenges**, including the inability to teach AI systems to think creatively and unconventionally, their limitation to primarily routine tasks, the minimization of direct human communication, and the weakening of incentives for personal and professional activity due to the reduced cognitive load on human neural processes.

6. Conclusion

The following key conclusions may be drawn from the conducted research:

1. As artificial intelligence continues to evolve, its influence on marketing activities will further intensify. From data analytics to the creation of creative content, companies will increasingly utilize AI-driven data for decision-making and the personalization of marketing campaigns.
2. Such personalization will extend beyond e-commerce and entertainment, and organizations should therefore be prepared to address challenges associated with algorithmic bias and data privacy concerns.
3. The use of artificial intelligence technologies requires maintaining a balance between the benefits provided by AI and protection against misuse or unethical practices.

As principal recommendations, international regulatory frameworks such as the **General Data Protection Regulation (GDPR)** — the European Union's comprehensive framework for personal data protection — and the **California Consumer Privacy Act (CCPA)**, which regulates the collection and use of personal data of California residents, may be considered as reference standards.

In order to remain competitive, companies should prioritize clear and ethical practices, including:

- informing consumers about how their data will be used;
- developing systems that minimize bias in AI-generated predictions;

- ensuring transparency in data processing practices and compliance with international data privacy legislation, including regulations such as the GDPR and CCPA.

Thus, artificial intelligence is transforming the ways in which companies predict and respond to consumer behavior, making AI an indispensable tool within the digital economy. By leveraging AI for hyper-personalization, churn prediction, and trend forecasting, organizations can gain deeper insights into customer preferences and create more meaningful interactions with consumers.

References

1. The Current State of Artificial Intelligence in Marketing in 2024. URL: https://ppc.world/articles/kak-iskusstvennyy-intellekt-menyaet-processy-v-marketinge-v-2024-godu/
2. SKURKO E.V. International Legal Regulation of Artificial Intelligence: the First Steps of International Organizations (Review article) DOI: 10.31249/iajpravo/2025.03.07 URL: https://cyberleninka.ru/article/n/mezhdunarodnoe-pravovoe-regulirovanie-iskusstvennogo-intellekta-pervye-shagi-mezhdunarodnyh-organizatsiy-obzornaya-statya/viewer
3. AI in Consumer Goods Production (Real-World Examples) URL: https://www.msrcosmos.com/ai-in-consumer-goods-industry-real-world-examples
4. Artificial Intelligence: Opportunities and Challenges for Moldova. URL: https://techdoor.md/iskusstvennyy-intellekt-vozmozhnosti-i-vyzovy-dlya-moldovy
5. AI Will Shape the Future of Marketing. URL: https://professional.dce.harvard.edu/blog/ai-will-shape-the-future-of-marketing/#The-Current-State-of-AI-in-Marketing
6. Artificial Intelligence Act, Official Journal version of 13 June 2024.
7. Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law. Council of Europe Treaty Series - No. 225. Vilnius, 5.IX.2024
8. Filina N. V., Gudkov Yu. A., Senin I. N. Legal Liability of Artificial Intelligence: Theoretical Models and Practical Solutions. Journal of Applied Research.2025. URL: https://cyberleninka.ru/article/n/yuridicheskaya-otvetstvennost-iskusstvennogo-intellekta-teoreticheskie-modeli-i-prakticheskie-resheniya/viewer
9. Innovations in Moldova: How AI Reduces Labor Shortages. URL: https://www.delucru.md/ru/articles/innovatsii-v-moldove-kak-ai-sokrashchaet-nekhvatku-rabochey-sili (10.06.2025)
10. For the First Time in Moldova, Artificial Intelligence Was Tested in the Creation of News Video Content. URL: https://noi.md/ru/nauka-i-it/v-moldove-vpervye-oprobovali-iskusstvennyj-intellekt-v-sozdanii-novostnogo-video

Comparative Analysis of International Digitalization Models for Agritourism Development

Svetlana Ceban

Academy of Economic Studies of Moldova, Chisinau, Moldova
ceban.svetlana@ase.md

Abstract: Digital transformation significantly influences the development of the tourism sector, including agritourism. The use of digital technologies such as online booking platforms, mobile applications, social media, and smart destination management systems contributes to increasing the visibility of rural tourist destinations and improving the tourist experience. The aim of this paper is to conduct a comparative analysis of international digitalization models in agritourism development. The research is based on comparative analysis and a review of the specialized literature, examining examples from Italy, Austria, Slovenia, and China. The results highlight the role of digitalization in online promotion, reservation optimization, and destination management, contributing to increased competitiveness of agritourism and to the sustainable development of rural communities.

Keywords: Digitalization, agritourism, rural tourism, digital transformation, international models

1. Introduction

In recent decades, rural tourism has become an increasingly important field within the tourism industry, contributing to the valorization of natural, cultural, and economic resources in rural areas. The development of tourism activities in rural environments creates new opportunities for economic diversification and for promoting local heritage. In the context of contemporary economic and social transformations, rural tourism contributes to the diversification of economic activities in rural areas and to increasing their attractiveness for tourists. At the same time, this type of tourism plays an important role in preserving the cultural and natural heritage of local communities and in promoting the sustainable development of rural areas [1].

The importance of rural tourism can be analyzed from multiple perspectives, including economic, social, and cultural aspects. From an economic point of view, rural tourism contributes to generating additional income for the local population, creating employment opportunities, and stimulating related economic activities such as traditional food production, handicrafts, and local services. Furthermore, the development of rural tourism contributes to the diversification of rural economies and reduces dependence on traditional agricultural activities, offering new development opportunities for local entrepreneurs and small and medium-sized enterprises in rural areas [2].

From a social and cultural perspective, rural tourism contributes to improving the quality of life in rural communities and reducing the phenomenon of rural depopulation. Through the development of tourism activities, the local population gains new economic and professional opportunities, which may reduce migration to urban areas and encourage young people to remain in rural regions. At the same time, rural tourism supports the preservation of

cultural heritage, the promotion of traditions, and the strengthening of the local identity of rural communities [2].

Within rural tourism, agritourism occupies an important place, representing a specific form of tourism based on the valorization of agricultural activities and rural lifestyles. By integrating agricultural activities with tourism services, agritourism contributes to the diversification of economic activities in rural areas and to the creation of additional sources of income for farmers, playing an important role in the development of the local economy [3].

In the context of globalization and digital transformation, the tourism sector is increasingly influenced by the use of information technologies and digital platforms. The development of digital platforms, mobile applications, and social media has significantly transformed the way tourists obtain information about destinations, plan travel experiences, and access tourism services. These digital tools facilitate direct interaction between service providers and tourists and contribute to increasing the accessibility of tourism offers. The use of such technologies also contributes to enhancing the visibility of tourist destinations and improving the overall tourist experience [4].

Under these conditions, digitalization becomes an essential factor for the modernization and increased competitiveness of the agritourism sector. However, the specialized literature highlights significant differences between the digitalization models implemented in different countries, which makes a comparative analysis of these international experiences necessary.

In this context, the analysis of international digitalization models in the development of agritourism becomes relevant for identifying best practices that can contribute to the sustainable development and increased competitiveness of rural tourism. Therefore, the contribution of this study consists in highlighting the main international digitalization models in agritourism and identifying the elements that can support the modernization of rural tourism.

The objective of the research is to identify and compare the main international digitalization models used in the development of agritourism.

The research hypothesis is based on the premise that the integration of digital technologies contributes to increasing competitiveness and to the modernization of agritourism destinations.

2. Theoretical Framework of Digitalization in Tourism and Agritourism

Digitalization can be understood as the process through which digital technologies are integrated into economic and organizational activities with the aim of improving process efficiency, facilitating access to information, and developing new forms of interaction between service providers and users. In the context of the digital economy, digitalization contributes to transforming the way organizations operate by influencing production, distribution, and communication processes with customers. In the tourism sector, digitalization represents an essential factor for adapting to the requirements of the global market and for increasing the competitiveness of tourist destinations [5].

Digital transformation in tourism involves the use of information technologies and digital tools to optimize the management, promotion, and commercialization processes of tourism services. The development of the Internet and mobile technologies has significantly changed tourist behavior, allowing tourists to quickly access information about destinations, accommodation services, and recreational activities. Currently, tourists frequently use digital platforms to compare tourism offers, make online reservations, and evaluate tourism experiences through reviews published on the Internet.

Digital technologies play an important role in promoting tourist destinations by facilitating rapid access to information and increasing the visibility of tourism offers on the international market. The use of digital tools enables tourism organizations and entrepreneurs in the tourism sector to develop effective promotion strategies and attract a larger number of visitors. In this context, online booking platforms represent one of the most widely used digital tools in the tourism sector, allowing tourists to compare different accommodation options and make reservations in a fast and accessible way. Among the most well-known platforms used globally are Booking.com, Airbnb, and TripAdvisor, which facilitate direct interaction between tourism service providers and tourists [6].

Another important element of digitalization in tourism is the use of mobile tourism applications, which provide tourists with access to information about tourist attractions, travel routes, local events, or services available within a particular destination. The use of mobile applications enables tourists to quickly access information about destinations, routes, or available services, thus contributing to a more informed tourist experience that is better adapted to individual preferences.

At the same time, digital marketing has become an essential tool for promoting tourist destinations and agritourism services. The use of social media allows tourism organizations to communicate directly with potential tourists and promote authentic tourism experiences through images, videos, and reviews. Platforms such as Facebook, Instagram, and TikTok are frequently used to promote tourist destinations and increase the visibility of tourism offers in the online environment.

Furthermore, the development of innovative technologies such as virtual reality and augmented reality provides new opportunities for promoting tourist destinations. Through these technologies, tourists can virtually explore certain destinations or tourist attractions before making the decision to travel. The use of such digital solutions contributes to increasing the attractiveness of tourist destinations and improving tourism marketing strategies [7].

In the agritourism sector, digitalization plays an increasingly important role in promoting rural tourism services and facilitating tourists' access to information about agritourism destinations. The integration of digital platforms, online marketing strategies, and innovative technologies can significantly contribute to the modernization of agritourism and to increasing its competitiveness on the international tourism market. Thus, digitalization represents an essential element for the sustainable development of rural tourism and for its adaptation to the new trends of the global tourism industry.

Based on the specialized literature, the main digital tools used in the development of agritourism can be identified and are presented schematically in Figure 1.

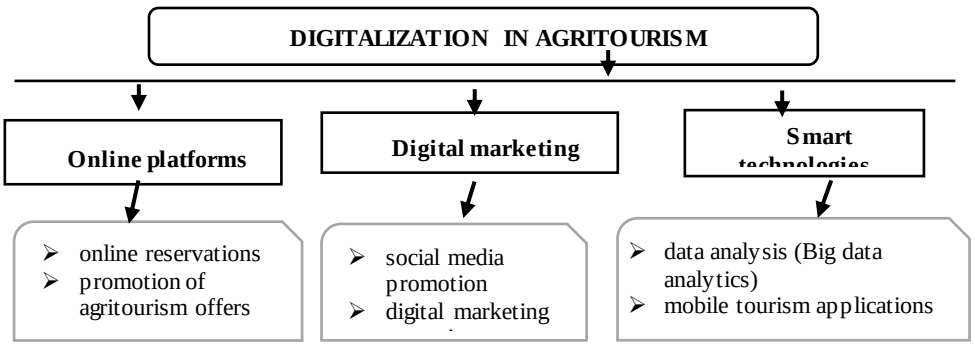


Figure 1. Conceptual model of digitalization in agritourism
Source: elaborated by the author based on the specialized literature.

The figure illustrates the main categories of digital technologies used in the development of agritourism, including online booking platforms, digital marketing tools, and smart technologies used in destination management.

3. Research Methodology

The present research is based on the use of the **comparative analysis method**, applied in order to identify, examine, and evaluate different international digitalization models implemented in the agritourism sector. This methodological approach enables the systematic comparison of digital transformation strategies adopted in various countries, allowing the identification of both similarities and differences in the way digital technologies are integrated into agritourism activities. Through this analysis, it becomes possible to highlight best practices and successful experiences that may contribute to the improvement and development of the agritourism sector in other regions.

The comparative analysis method is widely used in studies related to tourism development and digital transformation, as it facilitates a deeper understanding of how different economic, technological, and institutional environments influence the adoption of digital solutions in tourism. In addition, this method allows researchers to evaluate the effectiveness of digital tools and strategies in supporting the competitiveness and sustainability of rural tourism destinations [8].

The research also involved a **comprehensive review of the specialized literature**, including academic publications, scientific articles, conference papers, and international reports related to tourism development, digital transformation, and rural innovation. The literature review made it possible to identify the main theoretical approaches and conceptual frameworks associated with digitalization in tourism, as well as to highlight the increasing importance of digital technologies such as online platforms, mobile applications, and digital marketing tools in promoting rural tourist destinations and improving the accessibility of tourism services.

Furthermore, the research includes the **analysis of several international case studies** selected based on their relevance to the agritourism sector and their advanced level of digitalization. The analyzed case studies originate from countries that have successfully implemented innovative digitalization strategies in rural tourism, such as Italy, Austria, Slovenia, and China. These countries represent relevant examples due to their well-developed agritourism infrastructure and their active use of digital technologies to support tourism services, promotion, and customer interaction. The analysis of these case studies provides valuable insights into the practical implementation of digital solutions and allows the identification of best practice models that may contribute to the modernization and sustainable development of the agritourism sector.

For the comparative analysis, several **evaluation criteria** were considered. These include the level of integration of digital technologies in tourism services, the availability and use of online booking platforms for agritourism accommodations, the implementation of digital marketing strategies for the promotion of rural destinations, as well as the overall impact of these digital tools on the development and visibility of agritourism activities. Additional aspects such as the accessibility of digital information, the use of social media for tourism promotion, and the involvement of local communities in digital initiatives were also taken into account.

The analysis of these criteria allows for a better understanding of how digitalization contributes to increasing the competitiveness of rural tourist destinations, improving the efficiency of tourism management, and supporting the sustainable development of local communities. At the same time, the research highlights the potential benefits of adopting

modern digital solutions in agritourism, including enhanced visibility on international markets, improved communication with tourists, and the diversification of tourism services offered in rural areas.

4. International Digitalization Models in Agritourism

The digital transformation of the tourism sector has encouraged many countries to develop **comprehensive digital strategies and technological platforms** aimed at promoting rural tourist destinations and modernizing the agritourism sector. In recent years, the integration of digital technologies into tourism activities has become an essential component of tourism development policies, particularly in rural areas where tourism represents an important source of economic diversification. The analysis of international experiences highlights that the use of digital technologies significantly contributes to increasing the visibility of rural tourist destinations, improving the accessibility of tourism services, and strengthening communication between service providers and tourists. At the same time, digital solutions support the promotion of local traditions, gastronomy, and cultural heritage, which are key elements of the agritourism experience.

A relevant example in this regard is **Italy**, where agritourism represents a well-developed and economically significant sector of the tourism industry. The country has a long tradition in agritourism, supported by both governmental policies and private initiatives aimed at promoting rural tourism. In order to facilitate the promotion and accessibility of agritourism services, several digital platforms have been developed that connect tourists with agritourism farms located in different regions of the country. One example is the **Agriturismo.it** platform, which allows users to identify, compare, and book agritourism accommodations in various Italian regions. Through this platform, tourists can access detailed information about accommodation facilities, available activities, prices, and customer reviews, which significantly simplifies the decision-making process when choosing a rural tourism destination. At the same time, the platform contributes to increasing the online visibility of agritourism farms and facilitates direct interaction between service providers and potential visitors, thus supporting the sustainable development of rural tourism [9].

Austria also represents an important example in the digitalization of rural tourism and the promotion of agritourism activities through modern technological solutions. Numerous agritourism households and rural guesthouses use specialized online platforms and booking systems to promote their tourism services, present accommodation offers, and manage reservations efficiently. In addition, Austrian tourism authorities have developed advanced **digital destination management systems**, which integrate comprehensive information related to accommodation facilities, tourist attractions, local events, cultural activities, and available recreational services. These digital systems support both tourists and tourism operators by providing centralized and updated information about tourism opportunities within rural regions. As a result, digitalization contributes to improving tourism management, optimizing the promotion of rural destinations, and enhancing the overall tourist experience.

In **Slovenia**, the digitalization of rural tourism is closely associated with the innovative concept of **Tourism 4.0**, which promotes the use of advanced digital technologies for the development of smart and sustainable tourist destinations. The Tourism 4.0 concept is based on the integration of data analysis, digital platforms, artificial intelligence, and intelligent information systems in order to optimize destination management and improve tourism services. Through this approach, tourism stakeholders are able to collect and analyze large volumes of data related to tourist behavior, preferences, and mobility patterns. This information allows tourism providers to develop more personalized services and to design tourism experiences that better respond to the expectations of visitors. The implementation of

Tourism 4.0 also encourages collaboration between tourism businesses, local authorities, and technological innovators, contributing to the creation of an integrated and digitally supported tourism ecosystem[10].

Another relevant example is **China**, where the development of the **Smart Rural Tourism** concept has been supported by significant investments in digital infrastructure, information technologies, and the development of advanced online platforms dedicated to the promotion of rural tourist destinations. Within this model, digital technologies are widely used to promote rural attractions, facilitate online reservations, manage tourism flows, and provide tourists with real-time information about tourism services and activities. In addition, technologies such as mobile applications, digital maps, and intelligent recommendation systems are used to improve tourists' access to information and to enhance their travel experience. The implementation of these digital solutions contributes not only to increasing the attractiveness and accessibility of rural tourism destinations but also to stimulating local economic development by supporting small businesses, local producers, and rural communities involved in tourism activities [11].

The analysis of these international models highlights that digitalization plays an **essential role in the development and modernization of the agritourism sector**, as well as in increasing the competitiveness of rural tourist destinations at both national and international levels. The implementation of digital platforms, the use of online marketing strategies, and the integration of innovative technologies contribute to improving the promotion of agritourism services, optimizing tourism management processes, and creating more attractive tourism experiences for visitors. At the same time, digital transformation supports the sustainable development of rural areas by facilitating better access to markets, encouraging innovation, and strengthening the economic resilience of local communities.

5. Comparative Analysis of International Models

The comparative analysis of international digitalization models in agritourism allows the identification of the main strategies used to modernize the rural tourism sector and to increase the competitiveness of tourist destinations. The experiences of different countries demonstrate that the integration of digital technologies plays a key role in transforming traditional rural tourism into a more accessible, efficient, and market-oriented activity. Through the implementation of digital tools, tourism operators are able to promote their services more effectively, improve communication with tourists, and optimize operational processes. In this context, digital solutions such as online platforms, mobile applications, and intelligent destination management systems become essential instruments for the sustainable development of agritourism and for strengthening the visibility of rural tourist destinations on national and international markets [5].

In Italy, the development of agritourism is strongly supported by specialized digital platforms that enable the promotion of agritourism farms and the efficient management of online reservations. A relevant example is the Agriturismo.it platform, which operates as a digital marketplace dedicated to agritourism accommodations across different regions of Italy. This platform allows tourists to search for accommodation based on location, type of services offered, price range, and available activities, such as farm experiences, local gastronomy, or outdoor recreation. By providing detailed information, images, and user reviews, the platform facilitates informed decision-making for tourists and encourages direct interaction between visitors and agritourism providers. As a result, the visibility of agritourism farms increases significantly, while service providers gain improved access to broader tourism markets.

In Austria, the digitalization of rural tourism is supported by specialized platforms and networks dedicated to farm holidays, such as Farm Holidays in Austria. These platforms

contribute to the online promotion of rural accommodation units, offering comprehensive information about available services, accommodation facilities, and tourism activities in rural regions. In addition, these digital tools facilitate direct communication between hosts and tourists, allowing visitors to request additional information, verify availability, and make reservations online. Through these solutions, agritourism operators are able to strengthen their online presence and improve the promotion of authentic rural experiences, which contributes to increasing the attractiveness of rural destinations and stimulating local tourism development.

In Slovenia, the Tourism 4.0 model represents one of the most innovative initiatives aimed at integrating advanced digital technologies into the tourism sector. This concept promotes the use of data analysis, digital platforms, and intelligent systems in order to optimize tourism destination management and improve the quality of tourism services. Tourism 4.0 encourages the collection and analysis of data related to tourist behavior, preferences, and travel patterns, which allows tourism stakeholders to design personalized tourism experiences and improve service delivery. Moreover, this model promotes collaboration between tourism providers, technology developers, and public authorities, contributing to the creation of smart tourist destinations that are both efficient and sustainable.

In China, the concept of smart rural tourism is closely associated with the extensive use of digital infrastructure, mobile technologies, and intelligent management solutions aimed at revitalizing rural areas and increasing the revenues generated by rural tourism activities. Within this model, digital platforms and mobile applications are used to promote rural attractions, manage reservations, provide real-time information about tourism services, and facilitate communication between tourists and service providers. These technological solutions also help improve destination management and tourism planning by enabling better monitoring of tourism flows and visitor preferences. As a result, the implementation of digital technologies contributes to the modernization of tourist destinations, the diversification of tourism services, and the strengthening of the economic development of rural communities.

Based on these examples, the following comparative synthesis can be presented:

Table 1. International models of digitalization in agritourism

Country	**Digitalization model**	**Technologies used**	**Impact on agritourism**
Italy	Digital platforms for agritourism farms	online platforms, digital reservations	increased visibility of farms
Austria	Digitalization of rural tourism services	tourism applications, digital marketing	increase in the number of tourists
Slovenia	Tourism 4.0	big data, intelligent platforms	development and modernization of rural tourism
China	Smart rural tourism	mobile applications, intelligent systems	modernization of tourist destinations

The comparative analysis highlights that digitalization represents a key factor in the modernization and sustainable development of agritourism. The experiences analyzed demonstrate that the integration of digital technologies contributes not only to improving the visibility of rural tourist destinations but also to optimizing the management of tourism services and strengthening the relationship between tourism providers and visitors. By adopting digital solutions, agritourism operators are able to reach a wider audience, simplify

the process of accessing tourism services, and provide more detailed and accessible information about the experiences offered in rural areas.

Although the digitalization models implemented in different countries vary depending on economic conditions, technological infrastructure, and national tourism policies, they all emphasize the importance of digital technologies as essential tools for the development of modern tourism systems. In particular, digital platforms and online marketplaces play a significant role in promoting agritourism services, allowing tourists to easily search for destinations, compare offers, and make reservations through integrated booking systems. At the same time, the use of digital marketing strategies, social media platforms, and online communication tools facilitates direct interaction between tourism providers and potential visitors, contributing to the creation of a more dynamic and competitive tourism environment.

Furthermore, digital technologies support the more efficient management of tourism destinations by enabling the collection and analysis of tourism-related data, improving decision-making processes, and facilitating better coordination between stakeholders involved in tourism development. In this context, the integration of digital solutions such as online booking systems, mobile applications, and intelligent destination management platforms contributes significantly to increasing the efficiency, accessibility, and competitiveness of agritourism services. Therefore, digitalization can be considered an essential component of modern tourism development strategies, particularly in rural areas where tourism plays an important role in supporting local economies and promoting sustainable community development [7].

6. Conclusions

The conducted analysis highlights the important role of digitalization in the evolution and transformation of the tourism sector, including the development of agritourism activities. In the context of increasing technological progress and the growing demand for accessible and well-organized tourism services, the integration of digital technologies into rural tourism activities has become an essential factor in improving the efficiency of tourism operations. Digitalization contributes to optimizing key processes such as the promotion of tourist destinations, the management of reservations, and the coordination of tourism services, thus facilitating both the work of tourism providers and the experience of visitors.

The analysis of international models shows that the use of online booking platforms, mobile applications, digital marketing strategies, and intelligent destination management systems can have a significant impact on the development of agritourism. These digital tools enable tourism operators to promote their services more efficiently, reach a broader audience, and provide potential visitors with detailed and easily accessible information about rural tourism destinations. In addition, the implementation of digital solutions simplifies the reservation process and allows tourists to compare tourism offers, check availability in real time, and select services that correspond to their preferences. The examined international experiences demonstrate that the use of digital technologies enables more effective promotion of rural destinations and facilitates tourists' access to agritourism services, contributing to the growth of this tourism segment.

The comparative analysis also revealed differences between the digitalization models implemented in different countries. In some cases, the focus is primarily placed on the development of specialized digital platforms designed to promote agritourism services and facilitate online reservations. In other situations, digitalization is integrated into broader and more complex strategies aimed at developing smart tourist destinations based on the use of data analysis, intelligent information systems, and integrated digital infrastructures. These differences are influenced by several factors, including the level of technological development,

the availability of digital infrastructure, national tourism policies, and the degree of collaboration between public authorities and private tourism operators.

Therefore, one of the main conclusions of the research is that digitalization no longer represents only an optional element, but rather a necessity for the modernization and sustainable development of the agritourism sector. The adoption of digital technologies contributes to increasing the competitiveness of rural tourist destinations, attracting a larger number of visitors, and improving the overall quality of tourism services. At the same time, digitalization supports the diversification of tourism activities and enables rural tourism providers to better adapt to the expectations and needs of modern tourists.

In the future, the development of the agritourism sector will largely depend on the capacity of local stakeholders—such as tourism operators, rural entrepreneurs, and public authorities—to integrate digital solutions adapted to the specific characteristics of rural areas. These solutions may include the use of online promotion and booking platforms, communication and marketing through social media channels, the development of mobile applications dedicated to rural tourism, and the implementation of intelligent management tools for tourism destinations. The effective and coordinated use of such digital practices can contribute to the modernization of tourism services, improve the visibility of rural destinations, and support the sustainable development of local communities.

At the same time, it is important to acknowledge that this research presents certain limitations. These limitations are mainly related to the use of the comparative analysis method and the relatively limited number of case studies examined within the study. Although the selected examples provide valuable insights into the digitalization of agritourism in different countries, they do not fully capture the diversity of digital transformation processes occurring in rural tourism worldwide. Therefore, future research could expand the scope of the analysis by including a larger number of countries and regions, as well as by applying quantitative research methods. Such approaches would allow for a more detailed evaluation of the impact of digitalization on the development of agritourism and would provide more precise data regarding the economic, social, and technological effects of digital transformation in rural tourism.

References

[1] World Tourism Organization. (2020). Rural Tourism and Development. Madrid: UNWTO.
[2] Lane, B., & Kastenholz, E. (2015). Rural tourism: The evolution of practice and research approaches – towards a new generation concept? Journal of Sustainable Tourism.
[3] Phillip, S., Hunter, C., & Blackstock, K. (2010). A typology for defining agritourism. Tourism Management.
[4] Sharpley, R. (2002). Rural tourism and the challenge of tourism diversification. Tourism Management.
[5] Buhalis, D., & Law, R. (2008). Progress in information technology and tourism management: 20 years on and 10 years after the Internet. Tourism Management.
[6] Xiang, Z., & Gretzel, U. (2010). Role of social media in online travel information search. Tourism Management.
[7] Guttentag, D. (2010). Virtual reality: Applications and implications for tourism. Tourism Management.
[8] OECD. (2020). Tourism Trends and Policies. Paris: OECD Publishing.
[9] Agriturismo.it. (2024). Official platform for agritourism accommodation in Italy. Available at: https://www.agriturismo.it
[10] Tourism 4.0 Partnership. (2021). Tourism 4.0 – Data-driven tourism development. Available at: https://tourism4-0.org
[11] Zhang, Q., et al. (2022). Digitalization and rural tourism development in China. Sustainability.

Listening in English for IT Students: Challenges and Solutions

Anastasia Romanova

Academy of Economic Studies of Moldova, Chisinau, Moldova
romanova.anastasia@ase.md

Abstract. Listening comprehension is a fundamental skill in English for Specific Purposes (ESP), particularly for Information Technology (IT) students. Given the increasing role of English as the lingua franca of the digital and technological world, proficiency in listening is essential for accessing global resources, collaborating across borders, and engaging with rapidly evolving knowledge. However, IT students encounter specific challenges when developing listening competence in English. This paper outlines the main difficulties they face and proposes pedagogical solutions to enhance listening skills in the context of IT education. First, linguistic challenges stem from the fast pace of authentic speech, the use of technical terminology, and the frequent occurrence of accents from non-native speakers in the global IT environment. Second, cognitive challenges arise due to limited exposure to extended discourse in English, causing difficulties in processing information in real time. Third, psychological factors such as listening anxiety, fear of missing details, and low confidence can negatively impact comprehension. Additionally, the dominance of written communication in IT studies often leads to underdeveloped listening practice compared to reading and writing skills. To address these issues, a multi-layered approach is necessary. Task-based learning with authentic audio materials—such as conference talks, podcasts, and webinars—should be systematically integrated into curricula. Explicit training in IT-related vocabulary and discourse patterns can reduce the cognitive load. Incorporating interactive listening tasks (e.g., peer discussions based on video case studies, collaborative note-taking) fosters active engagement. Technology-enhanced learning tools, such as speech recognition software and adaptive listening platforms, provide personalized practice and immediate feedback. Finally, fostering metacognitive awareness by teaching listening strategies (predicting, inferencing, selective attention) helps students become more autonomous listeners. To conclude, listening in English presents unique challenges for IT students, yet these can be mitigated through targeted pedagogical interventions. By combining authentic exposure, task-based methods, and technological support, educators can equip future IT professionals with the listening competence required for successful participation in the global knowledge economy.

Keywords. Listening comprehension, English for IT, ESP, challenges, solutions, digital learning.

1. Introduction

In the context of globalization and rapid technological advancement, English has become the dominant language of information exchange, professional communication, and knowledge production in the field of information technology (IT). As a result, English language proficiency—particularly listening comprehension—has gained increased importance in higher education institutions that prepare specialists for the global digital economy. For IT students, the ability to understand spoken English is essential not only for academic success but also for effective participation in professional environments that rely heavily on international collaboration and continuous learning.

Listening comprehension is widely recognized as one of the most complex and cognitively demanding language skills. Unlike reading, listening requires real-time processing of information, interpretation of meaning, and adaptation to different accents, speech rates, and discourse styles. These challenges are especially pronounced in technical domains, where spoken input is often dense with specialized terminology and delivered at a rapid pace. Consequently, insufficient listening skills can limit students' access to academic content, reduce engagement in professional communication, and hinder long-term career development.

At the Academy of Economic Studies of Moldova (ASEM), IT students are increasingly exposed to English-language materials, including lectures, tutorials, webinars, and online courses. Despite this exposure, many students continue to experience difficulties in understanding authentic spoken English, particularly in professional and academic contexts. These difficulties point to the need for a deeper examination of the factors that affect listening comprehension and the pedagogical approaches used to support its development.

The present article aims to analyze the importance of English listening skills for IT students at ASEM, identify the main challenges they encounter in developing listening comprehension, and propose effective strategies to address these challenges. In addition, the article highlights the crucial role of teachers in facilitating listening development through targeted instruction, strategic use of technology, and the creation of supportive learning environments. By addressing both linguistic and affective dimensions of listening, this study seeks to contribute to the improvement of English language education for IT students and to support their successful integration into the global digital economy.

2. Importance of Listening Skills for IT Students

English listening proficiency constitutes a fundamental component of academic success and professional competence for students specializing in information technology (**Romanova,** 2017). In the context of globalization and rapid digital transformation, English functions as the dominant language of technological innovation, professional communication, and knowledge dissemination. As a result, the ability to comprehend spoken English is essential for effective participation in both academic and professional environments.

First, strong listening skills provide IT students with access to global educational and technical resources. The majority of programming tutorials, software documentation, technical webinars, and expert lectures are produced and delivered in English. Without adequate listening comprehension, students may experience difficulties in understanding complex explanations, following demonstrations, or applying newly acquired knowledge in practice. This limitation can significantly hinder the learning process and reduce academic performance.

Second, listening proficiency is crucial for professional communication in the IT sector. Modern IT professionals frequently work in international and multicultural teams, participate in virtual meetings, attend online conferences, and collaborate on cross-border projects. In such settings, English is commonly used as the primary working language, and effective listening is necessary to ensure accurate understanding of instructions, technical discussions, and collaborative decision-making.

Finally, continuous learning is an integral aspect of an IT career, as technologies and programming languages evolve rapidly. Online learning platforms such as Coursera, Udemy, and YouTube play a significant role in professional development; however, successful engagement with these resources depends largely on learners' ability to understand spoken English. Consequently, well-developed listening skills enable IT students to engage in lifelong learning and adapt to the changing demands of the global labor market.

Overall, the development of English listening competence enhances the academic achievement, professional readiness, and global competitiveness of Moldovan IT students in an increasingly interconnected digital economy.

3. Main Challenges Faced by ASEM IT Students

Despite the recognized importance of English listening skills for academic and professional success, IT students at the Academy of Economic Studies of Moldova (ASEM) encounter a range of challenges that hinder the effective development of listening comprehension. These challenges are both pedagogical and psychological in nature and are influenced by linguistic as well as contextual factors.

One of the primary difficulties is the **limited exposure to authentic English speech**. Many ASEM students acquire English predominantly through textbooks and structured classroom instruction, where speech is often simplified and carefully articulated. As a result, students frequently struggle to understand natural spoken English, including reduced forms, connected speech, regional accents, and informal expressions commonly used by native speakers. This gap becomes particularly evident when students are exposed to real-life materials such as academic lectures, podcasts, interviews, or conference presentations.

Another significant challenge relates to **technical vocabulary and rapid speech rate**. Audio materials in the field of information technology are typically dense with specialized terminology and are delivered at a fast pace. Although students may be familiar with technical terms in their written form, they often experience difficulty recognizing these terms in spoken discourse, especially when pronunciation differs from expected patterns or when speech is continuous and unsegmented. This mismatch between written knowledge and auditory recognition negatively affects comprehension and confidence.

A further obstacle is the **lack of focused and strategic listening practice**. In traditional educational settings, listening is frequently treated as a passive skill, with limited emphasis on the development of active listening strategies. Students may listen to recordings during lessons but rarely engage in structured activities that promote predicting content, identifying key ideas, distinguishing between main points and supporting details, or monitoring comprehension. The absence of systematic listening strategy training restricts students' ability to process complex spoken input effectively.

Psychological barriers also play an important role in limiting listening comprehension. Many students experience listening anxiety, particularly in academic or professional contexts where misunderstanding information may have negative consequences. Fear of missing important details or failing to understand the speaker can increase cognitive load, reduce concentration, and ultimately impair comprehension. Such affective factors are often underestimated but significantly influence listening performance.

Finally, the **influence of students' native languages**, primarily Romanian and Russian, presents additional challenges. Differences in phonetic systems, intonation patterns, rhythm, and stress between English and these languages can complicate the perception of unfamiliar sounds. This is especially problematic in cases of connected speech, weak forms, and idiomatic expressions, where meaning is not easily inferred from individual words. Consequently, linguistic interference may slow down processing and reduce overall listening accuracy.

Taken together, these challenges highlight the need for targeted pedagogical approaches that address both linguistic and affective dimensions of listening comprehension. Understanding these barriers is essential for designing effective instructional strategies aimed at improving English listening skills among ASEM IT students.

4. Solutions and Effective Strategies

Addressing the challenges associated with English listening comprehension among IT students requires a systematic and multidimensional approach that integrates pedagogical, technological, and psychological strategies. The following solutions are proposed to enhance listening skills among IT students at the Academy of Economic Studies of Moldova (ASEM) and to better align language instruction with professional and academic demands.

One effective strategy involves increasing **exposure to authentic listening materials**. Regular engagement with real-world English-language content—such as TED Talks, podcasts, professional IT webinars, and educational YouTube channels (e.g., Traversy Media or The Net Ninja)—allows students to become familiar with natural speech patterns, discourse structures, and field-specific terminology. Exposure to a range of English accents, including British, American, and Indian varieties, further contributes to the development of listening flexibility and improves overall comprehension in international communication contexts.

Equally important is the adoption of **active listening techniques**, which shift listening from a passive to an engaged cognitive process. Students should be encouraged to take notes while listening, summarize key ideas after listening tasks, predict content based on contextual cues, and listen selectively for specific information or keywords. These techniques help learners focus on meaning rather than attempting to understand every word, thereby reducing cognitive overload and improving comprehension efficiency.

The **systematic development of technical vocabulary** represents another critical component. Students can be encouraged to create personalized glossaries of IT-related terms, including phonetic transcription and pronunciation practice. Online tools such as YouGlish and Forvo are particularly useful for reinforcing accurate pronunciation and auditory recognition. In addition, integrating discipline-specific vocabulary into listening exercises enables learners to connect language acquisition directly with their field of study.

The **use of technology and online learning tools** further enhances listening instruction. Platforms such as Speechify, BBC Learning English, and ESLPod provide learners with opportunities to control playback speed, access transcripts, and engage in shadowing activities, where students repeat spoken language in real time. Moreover, AI-powered learning environments—including ChatGPT-based listening simulations—can offer interactive, adaptive, and personalized practice, addressing individual learner needs and proficiency levels.

Creating a supportive **English listening environment at ASEM** is also essential. Institutional initiatives such as IT-focused listening clubs, language cafés, guest lectures delivered in English by international experts, and collaborative projects with foreign universities can immerse students in authentic communicative situations. These activities extend learning beyond the classroom and reinforce the practical relevance of listening skills in academic and professional contexts.

Finally, attention must be given to **psychological readiness and confidence building**. Encouraging students to view errors as a natural part of the learning process can significantly reduce listening anxiety. Group-based activities, peer collaboration, and constructive feedback foster a supportive learning atmosphere, increasing students' confidence and motivation to engage actively in listening tasks.

Overall, the implementation of these strategies can contribute to the development of effective listening competence among ASEM IT students, enhancing their academic performance, professional preparedness, and competitiveness in the global digital economy.

5. The Role of Teachers in Promoting English Listening Skills for IT Students

Teachers play a central role in fostering the development of English listening skills among IT students, particularly in contexts where English functions as a language of professional communication and access to global knowledge. In higher education, instructors are not only providers of content but also facilitators of learning strategies, motivators, and designers of learning environments that support effective listening comprehension.

First, teachers are responsible for **selecting and integrating appropriate listening materials** that reflect authentic language use and professional relevance. For IT students, this involves incorporating discipline-specific audio and video resources such as technical lectures, webinars, interviews with industry professionals, and conference presentations. Research indicates that exposure to authentic and contextually relevant input enhances comprehension and learner motivation, especially when materials align with students' academic and career goals (Field, 2008; Vandergrift & Goh, 2012).

Second, teachers play a crucial role in **explicitly teaching listening strategies**. Listening comprehension does not develop automatically through exposure alone; it requires guided instruction in cognitive and metacognitive strategies such as predicting content, monitoring understanding, identifying key information, and reflecting on listening performance. According to Vandergrift and Goh (2012), strategy-based instruction significantly improves learners' ability to process spoken language and increases their autonomy as listeners. For IT students, such instruction is particularly important when dealing with fast-paced, information-dense technical discourse.

Another important aspect of the teacher's role is **scaffolding and task design**. Teachers can support learners by structuring listening activities into pre-listening, while-listening, and post-listening stages. Pre-listening activities activate prior knowledge and introduce key terminology, while post-listening tasks encourage analysis, discussion, and application of information. This structured approach reduces cognitive overload and facilitates deeper comprehension (Nation & Newton, 2009).

Teachers also contribute to listening development by **creating a supportive and low-anxiety learning environment**. Listening anxiety is a common barrier, especially when students fear misunderstanding or missing critical information. Instructors who encourage risk-taking, normalize errors, and provide constructive feedback help reduce affective barriers and promote learner confidence (Graham, 2017). Collaborative tasks and peer-based activities further support this process by allowing students to negotiate meaning in a less threatening context.

Finally, teachers act as **mediators of technology-enhanced listening practice**. By guiding students in the effective use of digital tools—such as platforms with adjustable playback speed, transcripts, and interactive exercises—teachers help learners develop independent listening skills beyond the classroom. In line with contemporary approaches to English for Specific Purposes (ESP), teachers must also continuously adapt their pedagogical practices to technological innovation and the evolving needs of the IT labor market (Hutchinson & Waters, 1987).

In summary, teachers play a pivotal role in promoting English listening skills for IT students through material selection, strategy instruction, pedagogical scaffolding, emotional support, and the integration of technology. Their active involvement is essential for preparing students to meet the linguistic and professional demands of the global digital economy.

6. Conclusion

Listening in English is both a challenge and an opportunity for IT students at the Academy of Economic Studies of Moldova. The main obstacles—limited exposure, fast speech, and technical vocabulary—can be overcome through consistent practice, strategic approaches, and the integration of authentic materials. As Moldova continues to strengthen its digital economy and international cooperation, the ability of ASEM graduates to **listen, understand, and respond effectively in English** will be a decisive factor in their professional success.

This article has examined the significance of English listening comprehension for IT students at the Academy of Economic Studies of Moldova, identifying key challenges that hinder effective listening development and proposing practical solutions to address them. In the context of globalization and rapid digital transformation, listening competence in English has emerged as a critical skill that underpins academic success, professional communication, and lifelong learning in the IT field.

The analysis demonstrates that ASEM IT students face a complex set of difficulties, including limited exposure to authentic spoken English, challenges related to technical vocabulary and fast speech, insufficient strategic listening practice, psychological barriers such as anxiety, and interference from native languages. These factors collectively reduce listening efficiency and confidence, particularly in professional and academic contexts that require comprehension of complex, information-dense spoken input.

The proposed solutions highlight the importance of a multidimensional approach to listening instruction. Increased exposure to authentic materials, systematic use of active listening strategies, targeted development of technical vocabulary, and the integration of digital and AI-based tools can significantly enhance listening comprehension. In addition, creating an immersive English listening environment at the institutional level and fostering students' psychological readiness play a crucial role in reducing anxiety and promoting sustained engagement. Furthermore, the article emphasizes the pivotal role of teachers in promoting listening competence. Through careful selection of relevant materials, explicit strategy instruction, structured task design, emotional support, and effective use of educational technologies, teachers act as key facilitators of listening development. Their role is particularly important in English for Specific Purposes (ESP) contexts, where listening skills must align closely with students' professional needs.

In conclusion, improving English listening skills among ASEM IT students requires coordinated efforts at both pedagogical and institutional levels. By addressing linguistic, cognitive, and affective dimensions of listening, higher education institutions can better prepare IT students to meet the demands of the global digital economy, enhance their employability, and support their long-term professional growth.

References

1. Field, J. (2008). Listening in the language classroom. Cambridge University Press.
2. Graham, S. (2017). Research into practice: Listening strategies in an instructed classroom setting. Language Teaching, 50(1), 107–119. https://doi.org/10.1017/S0261444816000306
3. Hutchinson, T., & Waters, A. (1987). English for specific purposes: A learning-centred approach. Cambridge University Press.
4. Nation, I. S. P., & Newton, J. (2009). Teaching ESL/EFL listening and speaking. Routledge.
5. Romanova, A. (2017). Listening comprehension facilitating strategies. *Euromentor Journal: Studies about Education, 8*(2), 65–75. ISSN 2068-780X.
6. Vandergrift, L., & Goh, C. C. M. (2012). Teaching and learning second language listening: Metacognition in action. Routledge

Decision Support Systems in Operational Risk Management: Trends, Perspectives and Threats

Olga Caminschi

Academy of Economic Studies of Moldova, Chisinau, Moldova
caminschi@ase.md

Abstract. Operational risk management is increasingly driven by the rapid advancement of information technology and the growing complexity of business processes. In this context, decision support systems (DSS) are becoming a key tool for improving the quality of management decisions through the integration of data analytics, scenario modeling, and artificial intelligence. Current trends demonstrate a shift from static reporting to dynamic predictive systems that provide real-time monitoring and proactive risk mitigation. However, the implementation of DSS is also associated with a number of threats, including data quality issues, cybersecurity vulnerabilities, overreliance on automated models, and the high cost of system integration. The purpose of this study is to analyze the dual nature of DSS in operational risk management, emphasizing their role as both a source of new opportunities for improving decision-making efficiency and a potential source of new risks.

Keywords. Operational risk, decision support systems, business continuity, cybersecurity.

1. Introduction

The current context of digitalization of the environment in which organizations operate is clearly shifting the focus of operational risk management toward the use of automated tools for risk identification and management. With the use of Big Data and the Internet of Things, and the availability of a large number of metrics and indicators that enable system monitoring and other risk indicators, Decision Support systems (DSS) are used to perform predictive analysis and make timely decisions. The purpose of this study is to determine how decision-making systems can optimize the operational risk management process and identify existing risks generated by this approach. Furthermore, to identify trends in the use of decision-making systems and develop recommendations for the application of such systems for effective risk management. In this context, the following objectives were set:

1. Determine the positive impact and significance of using these systems in risk management.
2. Identify trends in the use of operational risk management decision-making systems.
3. Identify risks that may arise from using these systems.
4. Develop recommendations for minimizing negative consequences.

Modern approaches to operational risk management involve the use of a range of tools, including risk identification and assessment, risk and control self-assessment (RCSA) procedures, key risk indicator (KRI) monitoring, operational event and loss analysis, and scenario analysis and stress testing. However, with the ever-increasing volume of data and the increasing complexity of organizational processes, traditional methods of information analysis and processing are becoming ineffective for the timely identification and assessment of risks.

The key research questions in this context are:

RQ1: Does the use of DSS have a positive impact on the risk management system?

RQ2: What negative consequences can the use of decision-making systems entail?

RQ3: How can the negative impact of a tool that is inherently designed to reduce risks be avoided?

Integrating DSS into operational risk management processes opens up new opportunities for improving risk management effectiveness. The use of analytical models, big data processing tools, forecasting methods, and information visualization enables the timely identification of trends, analysis of the causes of operational events, and assessment of the potential consequences of various risk scenarios. Thus, DSS can become an important element in the digital transformation of risk management systems, facilitating the transition from reactive responses to events to more proactive and analytically sound risk management. However, there are a number of risks that need to be considered when using these systems.

2. Theoretical and methodological foundations of DSS

Decision Support Systems (DSS) represent an important field within information systems research that focuses on improving managerial decision-making through the use of analytical models, data management technologies, and interactive software tools. According to Power (2002), DSS are computer-based information systems designed to support decision-makers in solving semi-structured and unstructured problems by integrating data, models, and user interfaces. The concept of DSS emerged in the 1960s and 1970s as a result of research on organizational decision-making and interactive computing systems. Early studies focused on using computer technologies to support managerial decision processes by providing structured access to organizational data and analytical models. These developments laid the foundation for modern DSS that integrate data management, modelling techniques, and analytical tools to support complex decision-making processes in organizations (Turban et al., 2011). Over time, DSS have evolved significantly alongside advances in information technologies. Early DSS were primarily model-driven systems designed to support quantitative decision analysis. Later developments incorporated database technologies, knowledge-based systems, and artificial intelligence techniques, transforming DSS into advanced analytical platforms capable of supporting both operational and strategic decision-making (Power, 2002; Turban et al., 2011). According to Arnott and Pervan (2008), DSS research focuses on the development and application of information systems designed to support managerial decision-making through analytical models, data processing techniques, and knowledge-based approaches. Such systems assist decision-makers in identifying problems, evaluating alternative solutions, and selecting optimal decisions.The main distinguishing feature of the use of DSS is the absence of subjectivity of the expert making the decision, so that the decision is made solely on the basis of achieving certain triggers or threshold values of pre-established parameters.

In order for the system to have the necessary parameters for formulating an informed decision that will subsequently be made, a certain sequence of actions or system configuration is required, which are determined by the following main stages:

1. Data collection, which feeds the system with a diverse set of information from document databases and other sources that enable the system's knowledge base to be populated in real time.
2. Data processing and modeling – enables data processing using various statistical, mathematical, and other algorithms to identify patterns and potential outcomes.
3. Analysis of processed data based on various scenarios. At this stage, the obtained information is processed in the context of decision-making in accordance with a

variety of scenarios. This action allows one to determine how other interrelated parameters will behave if a specific decision is made based on the available metrics.

4. Creation of a visual control board, which allows the existing trends and rationale for the system's decision-making to be easily visualized.

Figure 1 shows the components of the DSS system.

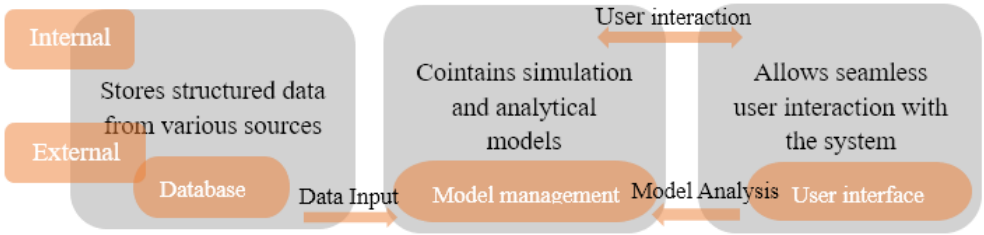


Figure 1. Components of the Decision Support System.
Source: developed by the author based on information from source [1, 2]

There are various types of decision support and decision-making systems. All are data-driven, utilizing data from various data stores to answer specific queries. Examples of such DSS include computer databases that contain queries for the system to validate. These systems are deployed via the web or a client-server architecture. A power system provides document-based decision support. These systems search websites and documents using a specific set of keywords or search queries. These systems are based on web technologies or client-server systems. There are also knowledge-based decision support systems. This broad range of systems is used by users both within an organization and as individuals interacting with the organization, such as a user. Complex systems that help analyze large data sets and make decisions or choose between options are categorized as model-based decision support systems. They can be used by both company employees and individuals interacting with the organization, depending on the goals of the system.

Analyzing the operational risk management system in the financial sector, the most strictly regulated sector in terms of risk management system development, it was revealed that the operational risk management process, particularly the collection, processing, and decision-making based on available data, has undergone changes over the past decades. Several developments have influenced the development of requirements or principles upon which decision-making systems should be based. The figure 2 presents the key developments that have impacted the operational risk management system.

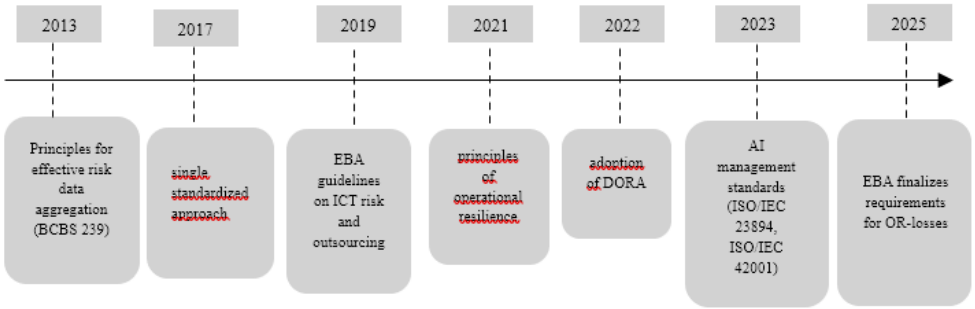


Figure 2. Key events that have impacted the operational risk management system
Source: developed by the author based on information from source [3, 4]

Thus, based on the presented figure, it becomes clear that the parametrization of decision-support systems needs to shift from static information recording to proactive responses based on available data. In modern organizations, changes occur rapidly, necessitating a shift from static, reactive responses to real-time, dynamic system management.

3. Integration of the Decision support system into operational risk management tools

Before analyzing methods for integrating a DSS, it should be noted that within the context of operational risk management, such a system is a tool for improving the quality, speed, and validity of decisions, not a substitute for expert judgment. The main stages of operational risk management are risk assessment, monitoring, control, and response as is shown on Figure 3.

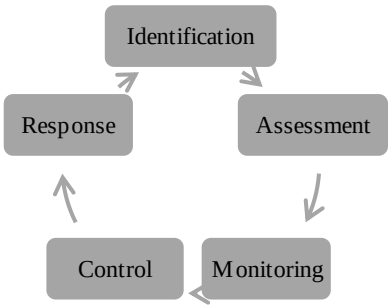


Figure 3. The main stages of operational risk management
Source: developed by the author based on information from source [3]

Within each of these stages, a DSS can provide support or an analytical basis for decision-making. Given that operational risk is generated by various factors, such as people, processes, systems, and external events, the need to process large amounts of data becomes clear. Furthermore, the parameters that determine compliance with certain acceptable thresholds are often located in different systems. Since operational risk management is a complex system encompassing various areas, from human resource management operations and customer complaints to purely technical transaction processing or other operational procedures, a unified system capable of integrating and processing data from various sources is essential. Given that operational risk management decisions are based not only on statistical analysis of historical data but also on forecasting, assessing possible scenarios, comparing various scenarios, and analyzing existing alternative solutions, the need for a forecasting model in managing this type of risk is obvious. One of the main advantages of using such systems is the identification of hidden dependencies and trends that cannot be determined through expert analysis. Furthermore, this approach provides a unified information context for decision-making by aggregating and structuring data. Given the above, a decision support system is an important component of effective risk management.

One of the operational risk management tools is self-assessment of risks and controls. This tool is based on the principle of risk assessment directly by process owners, as business process owners have the most accurate information about threats. The core concept of risk and control self-assessment is that business process owners describe their subprocesses to structurally identify existing risks, assess them based on their probability of occurrence and impact, and assess existing control measures. After the risk level is determined based on a

predefined impact and frequency assessment matrix, it is subjected to the existing control systemto determine the residual risk level.

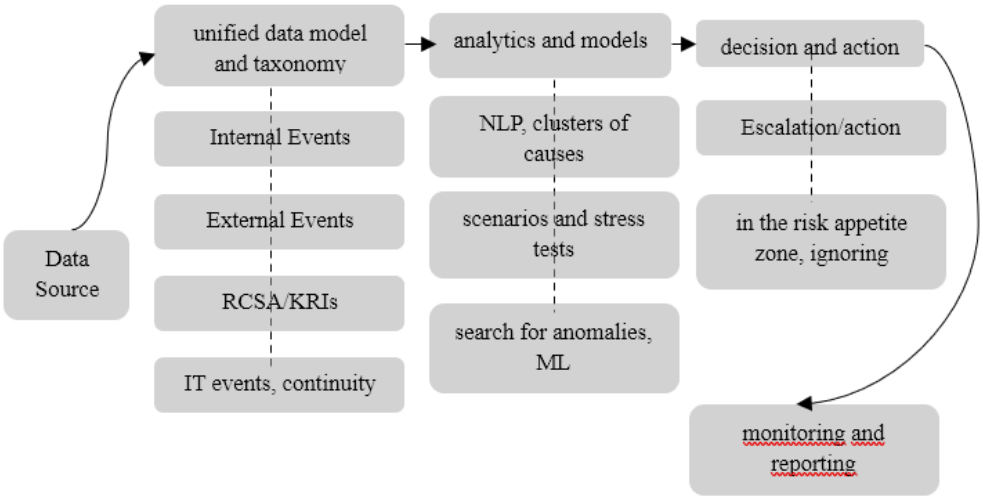


Figure 4. Integration of the Decision support systeminto operational risk management tools and process
Source: developed by the author based on information from source [2]

This tool allows for the identification of vulnerabilities in business processes before they lead to actual financial losses. Furthermore, this tool allows for the optimization of controls, i.e. determining whether measures are insufficient or excessive. This, in turn, is necessary for conducting an analysis taking into account the "cost-benefit-risk" principle, which allows for assessing the costs of maintaining controls for specific processes and the losses in the event of an operational risk event. It's worth noting that assessing control effectiveness should be considered in accordance with two key parameters: control design, or how logically and comprehensively the control's intended functioning is defined, and its operational effectiveness, or the extent to which employees actually implement the control.

Conducting a self-assessment of risks and existing controls consists of several stages. During the preparation stage, business processes are defined, followed by risk identification. When compiling a list of threats inherent to a given process, it is possible to integrate DSS into the process. When assessing the probability and potential impact of a risk, the systemenables comparison of expert assessments with historical data on actual incidents. This, in turn, prevents subjectivity in the analysis. Thanks to its ability to process large volumes of data, the systemcan take into account the frequency of events in similar processes, thereby facilitating the coverage of a larger number of processes during the self-assessment. System enables the identification of discrepancies between the formal assessments of process owners, which may be subjective, and actual violation statistics. For example, if a department assesses the risk of system unavailability as low, but statistics for previous periods indicate frequent incidents of this type, the system may indicate the need for additional analysis. This approach also applies to control assessment: if controls are formally declared, but incidents in this area recur periodically, the systemconsiders these controls ineffective, despite their inclusion in internal regulations. Because the system operates based on expert-defined parameters and is capable of self-learning, it is possible to use weighting factors, threshold values for various parameters,

and develop different scenarios for automated risk calculations in various processes. This approach is especially important for large enterprises, where a standardized approach to assessing various departments is required. Thus, the DSS not only enables expert assessment of the inherent risks and control levels in those processes by their owners but also makes this tool more evidence-based.

One of the focuses of the risk management system is identifying negative trends before the risk is converted into actual losses. For this purpose, various metrics or parameters are used at the enterprise level to measure the likelihood of an adverse event and assess the impact of its consequences. Key risk indicators (KRIs) are such metrics within operational risk management. These indicators provide early warning of increasing risk levels in various areas throughout the organization. When defining KRIs, it is necessary to evaluate the effectiveness of monitoring these indicators, specifically, the decisions that will be made based on the established metrics. The key functions of KRIs are to inform the risk management specialist when a parameter exceeds predetermined thresholds at various levels. Furthermore, these indicators can be used for forecasting based on the identified correlation between indicator growth and the likelihood of an incident. For example, if the current personnel indicator in a department with an increasing volume of operations significantly increases, the increased likelihood of an operational risk event in that process becomes apparent. This, in turn, allows one to justify the need for monitoring or allocating additional resources as a proactive response. There are different categories of risk indicators, depending on their purpose. For example, there are workload indicators that measure the volume of an operation, which increases the likelihood of error. Other indicators evaluate the quality of the internal environment or, for example, the level of employee training. Furthermore, there are indicators that report system errors. It's worth noting that the effectiveness of a key risk indicator system depends on the correctness of their initial setup, the definition of thresholds for various levels, as well as the quality of their monitoring and timely response. In this context, a decision support system acts as a tool capable of processing large amounts of data based on specified parameters. When operates, the system can collect data from various sources, track its dynamics in real time, and, accordingly, compare this data with the actual thresholds set. If negative trends are detected, the issue can be escalated for timely decision-making. For large organizations with various types of operations, products, and services, as well as various systems and technological infrastructure that support these operations, there is a need for simplified visualization of key risk indicator information in the form of dashboards and diagrams to generate analytical conclusions regarding metrics. The DSS allows for recording the exceeding of a threshold for a key risk indicator, as well as analyzing its relationship with another indicator and determining the level of impact of this exceeding.

Thus, there are a number of ways to integrate DSS into existing operational risk management tools.

4. Risks of Using the Decision Support Systems

When using DSS organizations face a number of risks or challenges during their implementation and integration. One of the most significant risks is model risk, which arises when using various algorithms—model-based, statistical, predictive, and decision-making—based on the results. Since DSS are often model-driven, meaning they utilize various models for decision-making, typical models used in these systems are distinguished. These include optimization models, simulation models, forecasting models, and what-if models. There are also data-driven models, which rely on big data processing and analytics. These models utilize machine learning, statistical analysis, and predictive analytics. The primary resources for these models are the database and its storage system, as well as the relationships between values

within the database. An example of using such models is identifying risk factors based on the analysis of historical operating losses. There are also knowledge-based models that use expert rules based on a knowledge. These models include expert systems, rule-based systems that automatically suggest actions when trends are detected. These systems rely not only on the use and processing of data but also on expert knowledge and established rules. Document-oriented models are another example of this type of model. These models work with unstructured information, meaning they can systematically search for documents, analyze texts, and analyze, for example, regulations, procedures, and reports for decision-making.

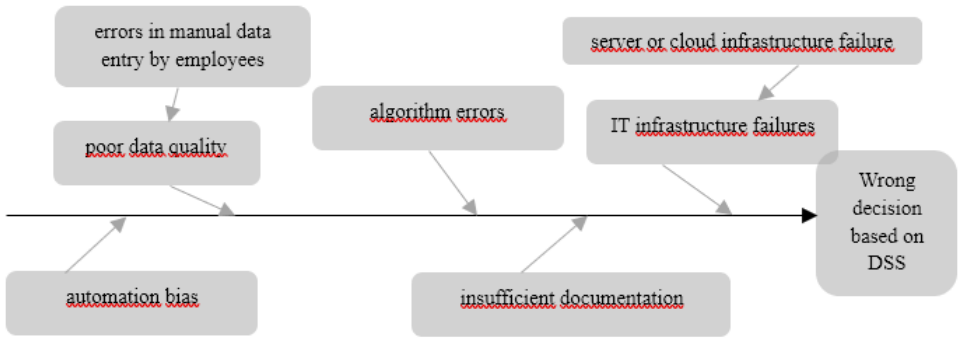


Figure 5. Cause and effect diagram of risk factors for using DSS in operational risk management
Source: developed by the author based on [1, 2, 3, 4]

Incorrect selection of the model type, incorrect calibration, and use of incorrect assumptions can lead to errors in algorithm development and incorrect statistical assumptions. Furthermore, the model must be regularly re-evaluated to determine its relevance in a dynamic external environment. Model risks associated with using a decision support system can lead to incorrect assessment of risk parameters (i.e., the probability of their impact), distorted indicators, and, consequently, incorrect management decisions. This, in turn, increases the organization's risk profile. Another important risk associated with using a decision support system is data quality. Since source data directly affects the accuracy of the decision support system's results, their quality plays a significant role in the system's effective use. In situations where data is incomplete or outdated, the system may formulate incorrect analytical conclusions and develop recommendations that are inconsistent with reality or do not mitigate the identified risk. The main challenges in ensuring data quality include errors in manual data entry, duplicate entries, lack of integration between various information systems that record data, and missing values in the data accounting system. This leads to the misidentification of high-risk areas, for example, when monitoring key risk indicators. A DSS is part of an organization's IT infrastructure; its availability directly depends on the reliability of its information technology. Therefore, when using these systems, technological risks are evident, including basic system availability risks: server or cloud infrastructure failure, software malfunctions, update errors, and incompatibility of new versions with other systems. Furthermore, there are challenges integrating DSS with other information systems to ensure the completeness and comprehensiveness of the analyzed data. Since decision-making systems, as noted above, process large amounts of sensitive data, often constituting

confidential information, these systems become potential targets for cyber-attacks (Caminschi O., 2025). This risk manifests itself in potential unauthorized access to the system, leakage of confidential information, external attacks on the data infrastructure, and the threat of malware injection. If these risks materialize, high-impact consequences may result, including the loss of confidential data or the distortion of analytical information used in risk management. Furthermore, data manipulation or modification by third parties may lead to erroneous decisions or the creation of an inaccurate perception of the risk level.

Given the current trend toward artificial intelligence and machine learning, decision-making systems can be difficult to interpret. In such situations, users may not understand how the system reached a certain conclusion, resulting in a lack of transparency in the algorithms used. This leads to a number of challenges, including the inability to justify decisions to regulators, difficulties auditing models, and difficulties interpreting results. The use of decision-making systems also poses the risk of overreliance on the system. This risk manifests itself through user overreliance on system recommendations, where employees perceive the recommendations and conclusions generated by the system as objective and requiring no expert review. This, in turn, leads to a decrease in critical thinking among employees, ignoring alternative approaches, and underestimating new or non-standard risks that fall outside the established parameters.

Thus, a number of potential risks are noticeable when using the Decision Support Systems.

5. Conclusions and Recommendations

With regard to RQ1 study's results demonstrate that integrating decision support systems significantly improves the effectiveness of risk management processes. Decision support systems improve the quality and speed of decision-making by enabling the systematic analysis of large volumes of data, the application of analytical models, and the visualization of risk-related information. However, the study also highlights several potential challenges associated with the use of such systems. In response to RQ2, the results suggest that the implementation of DSS may introduce new types of operational risks. These include risks related to poor data quality, algorithmic errors, technological failures, and excessive reliance on automated decision recommendations.

Addressing the RQ3, organizations are recommended to validate the models used before their integration, as well as periodically calibrate the algorithms, which is itself part of the model risk management system. Conducting independent checks of the algorithms of the models used will allow for the timely identification of existing deviations, confirm the relevance of the model, and determine the appropriateness of using the models in a given context.

To minimize data quality risks, companies must implement data quality management procedures. These procedures should ensure the definition of data standards, the presence of various verification processes, and the implementation of automated data cleansing mechanisms. To this end, companies should implement automated micro-algorithms for verifying the correctness of the entered data at each stage of the data entry process, generating notifications for the individuals entering this data. In this context, it is also necessary to conduct regular audits of the quality of the information used to confirm its relevance. It is worth noting that this risk is particularly pronounced in the context of using decision-making systems, and its management is fundamental.

To ensure the continued use of systems, it is necessary to manage IT risk. This requires implementing data backup measures, testing updates, and implementing IT infrastructure monitoring systems. In this regard, companies should implement a system availability

monitoring system that generates and sends notifications to responsible persons so that they can take action in the event of system unavailability. As part of cyber risk management, companies are recommended to employ modern cybersecurity mechanisms such as data encryption, intrusion detection, regular security testing and vulnerability scanning, and information system audits. Access control and access segregation are also important tools. The distribution of roles in the decision-making system also ensures controlled use or consumption of data accessible within the system.

To avoid the risk of insufficient algorithm transparency when implementing decision-making system models, it is recommended to use explainable artificial intelligence methods and ensure full, formal documentation of the model's logic. This requires developing procedures and internal regulations that demonstrate the logic and consistency of judgments used in making decisions. The logic of the algorithms used should be communicated to management so that they understand the methods of analyzing and processing the data used to make decisions. It is worth noting that decision-making by a DSS represents a transfer or delegation of competencies within the approved risk appetite; therefore, the governing body must carefully consider the approval of risk rules or algorithms. It is also worth noting that decision-making systems should be implemented in stages, using pilot testing and involving risk management specialists in the process of directly developing and configuring the rules upon which the DSS formulates conclusions and recommendations.

Thus, the implementation of DSS in operational risk management has a positive effect when the risks associated with their use are properly managed. A control system that includes data quality management, model validation, cybersecurity measures, and algorithm and model regulations will enable the effective use of these systems. Following the above recommendations will maximize the system's effectiveness in risk management.

References

[1] Power, D.J. 2002. Decision Support Systems: Concepts and Resources for Managers. Westport, CT: Greenwood Publishing Group.
[2] Turban, E., Sharda, R., Delen, D. and King, D. 2011. Decision Support and Business Intelligence Systems. 9th ed. Upper Saddle River, NJ: Pearson Education.
[3] Basel Committee on Banking Supervision. 2011. Principles for the Sound Management of Operational Risk. Basel: Bank for International Settlements.
[4] Basel Committee on Banking Supervision. 2014. Supervisory Guidelines for the Advanced Measurement Approaches for Operational Risk. Basel: Bank for International Settlements.
[5] Moeller, R. 2011. COSO Enterprise Risk Management: Establishing Effective Governance, Risk, and Compliance Processes. Hoboken, NJ: Wiley.
[6] Power, M. 2007. Organized Uncertainty: Designing a World of Risk Management. Oxford: Oxford University Press.
[7] Alter, S. 2004. 'A Work System View of DSS in Its Fourth Decade'. Decision Support Systems.
[8] Arnott, D. and Pervan, G. 2008. 'Eight Key Issues for the Decision Support Systems Discipline'
[9] Shim, J.P., Warkentin, M., Courtney, J.F., Power, D.J., Sharda, R. and Carlsson, C. 2002. Past, Present, and Future of Decision Support Technology'.
[10] Courtney, J.F. 2001. 'Decision Making and Knowledge Management in Inquiring Organizations: Toward a New Decision-Making Paradigm for DSS'. Decision Support Systems
[11] Caminschi, O. 2025. 'Cybersecurity as a Part of Operational Risk Management'. In: Proceedings of the International Scientific Conference "Development Through Research and Innovation – 2024", 4th Edition.

The Circular Supply Chain Model in Reducing Food Waste and Reusing Resources in the Food Industry

Eugeniu Stoian

Academy of Economic Studies of Moldova, Chisinau, Moldova
stoian@ase.md

Abstract. This article examines the role of the circular supply chain model in reducing food loss and increasing resource efficiency in the food industry. The relevance of this topic stems from the global scale of the food waste problem, which has a significant impact on food security, environmental sustainability, and the economic efficiency of agri-food systems. The aim of the research is to provide a theoretical foundation and practical analysis of the mechanisms for implementing the principles of the circular economy in the management of food supply chains. The paper reveals the essence of the circular economy as an alternative to the linear "production-consumption-disposal" model, as well as the main types of circular business models, including resource regeneration models, product life cycle extension, shared use of platforms, and the shift from product to service. Special attention is given to the structural stages of the food supply chain-from agricultural production and processing to logistics and retail-with a focus on the factors that cause losses and the possibilities for minimizing them. The methodological framework of the study includes a review of the literature, a comparative analysis of circular models, and an assessment of the impact of digital technologies (IoT, blockchain, data analytics) on the transparency and sustainability of supply chains. The results show that the implementation of circular practices contributes to waste reduction, cost savings, the strengthening of supply chain sustainability, and increased consumer confidence. This research was conducted as part of the RO-MD bilateral project, in partnership with the "ERNEST LUPAN" Institute for Research in Circular Economy and Environment (IRCEM), project code PN-ÎV-P8-8. 3-RO MD -2023- 0215 Studies and Investigations on the Interplay between Regenerative Agriculture and the Circular Economy in Romania and the Republic of Moldova.

Keywords: Circular supply chain, circular economy, food waste, food industry, resource recovery.

1. Introduction

Food waste is a serious global problem with significant implications for food security, resource use, and greenhouse gas emissions. According to the UNEP Food Waste Index Report of 2024, in 2022, approximately 1.05 billion tons of food waste were generated in retail, food services, and households-equivalent to 19% of the food available to consumers and over a billion wasted meals every day. This waste accounts for 8–10% of global anthropogenic greenhouse gas emissions and also contributes to chronic hunger affecting approximately 783 million people.[13]

In addition to wasteful consumption, the FAO estimates that approximately 13% of food is lost along the supply chain (from harvest to retail, excluding final consumption), highlighting the limitations of the traditional linear economic model.[14]

In the Republic of Moldova, food waste remains a persistent problem, characterized by high per capita rates in the Eastern European region, exacerbated by fragmented agricultural

supply chains, dependence on imports, and external vulnerabilities (such as climate variability). National data included in official statistical reports indicate a significant prevalence of losses and waste in the agricultural and consumer sectors, while UNEP's global reports do not include the most recent consolidated national estimates.[15]

The transition to a circular economy through the implementation of circular supply chains offers a structured solution: models such as resource recovery, extending product lifespans, and replacing primary resources with renewable or recycled resources enable the transformation of waste into productive resources and the reduction of systemic losses.[16].

2. Theoretical Aspects

In recent years, the concept of the circular economy has become one of the key theoretical approaches used to analyze sustainable development and the rational use of resources. In the scientific literature, the circular economy is considered an economic model focused on maximizing the longevity of products, materials, and resources in the economic cycle through processes of reuse, repair, recycling, and resource recovery.[11] Unlike the traditional linear economic model based on the "take-make-dispose" principle, the circular economy involves creating a regenerative system in which waste generation is minimized and resources are reused repeatedly.

The implementation of circular economy principles at the organizational level is achieved through the adoption of circular business models. According to research conducted by De Angelis (2018)[12], circular business models are models of economic activity in which customer base growth is achieved through offerings characterized by environmental efficiency, sustainability, and the rational use of resources. Such models allow organizations to more effectively control resource flows throughout the value chain and encourage cooperation among the various participants in economic processes.

The food supply chain is a complex system comprising several interconnected processes that, in unique ways, work together to move food from its source to its final destination. It all begins with agriculture, where livestock is raised and crops are grown. Farmers' products pass through processing centers, which prepare the products for consumption. Next comes the distribution stage, where processed, ready-to-eat foods are transported via logistics systems to wholesale or retail markets. Finally, the last stage of this process allows consumers to purchase food products from supermarkets, retail stores, or e-commerce sites. Throughout all these stages, food is produced, and technology is essential for improving production rates, streamlining logistics, and enhancing supply chain traceability to manage durable consumer goods and minimize waste.

3. Stages of the Food Supply Chain

1). Agriculture: Producers use modern methods and biotechnology to increase crop and livestock production based on climate and soil conditions.
2). Processing: Companies are introducing automation and artificial intelligence to improve the processes of transforming raw materials into processed products, in accordance with quality and safety standards.
3). Distribution: Supply chain software and IoT devices enable real-time visibility and optimal routing, which reduces delivery times and improves the efficiency of food transportation.
4). Retail: E-commerce systems and point-of-sale systems enable proper inventory control, while business analytics provide relevant insights that guide the personalization of the shopping experience.

4. Key Players Along the Supply Chain

1). Farmers and agricultural workers: As the name suggests, these people grow crops and raise livestock, using progressive farming methods to increase production.
2). Food processing companies: Companies that transform raw food materials into edible products for consumption, using their resources to maintain cleanliness and safety.
3). Logistics providers: Companies responsible for transporting and storing food products, utilizing the resources necessary to implement a supply chain.
4). Wholesalers and distributors: They handle bulk purchases of goods supplied by manufacturers or processors for resale to consumers through retailers, typically in large quantities to avoid running out of stock.
5). Retailers: Supermarkets, grocery chains, and online grocery sites that sell food products to end consumers, facilitated by technology for easy shopping.
6). Regulatory agencies: Government agencies that establish health and epidemiological requirements and controls regarding food safety and regulatory compliance.
7). Consumers: the final link in the food supply chain who ultimately purchase the products available on the market, thereby sustaining the supply chain.[2]

5. Research Methodology

The transition to a circular economy essentially involves designing and implementing new business models aimed at extending product lifespans and keeping resources in circulation through reuse, repair, recycling, recovery, and other restorative processes. Roberta de Angelis, author of the book "Business Models in the Circular Economy," defines circular models as business models in which customer growth is driven by comprehensive circular offerings, such as environmental performance, greater convenience, and superior durability. The new business models would enable.[9]

- ✓ Greater control over resource flows throughout the value chain;
- ✓ Innovation across the supply chain, paving the way for new business opportunities;
- ✓ Increased collaboration within the supply chain among relevant stakeholders;
- ✓ The creation of services that incorporate valuable products and resources.

The guide developed by the Forum for the Future identifies, describes, and examines case studies for 10 types of circular models, as follows: closed-loop recycling (this business model involves using waste as raw material to create new products), product-to-service (servitized products), customization of products or services (creates data management opportunities that enable product customization), dowcycling (a type of recycling in which the value of the new product obtained is lower than that of the original product), upcycling (a type of recycling in which the value of the new product obtained is higher than that of the original product), consumer loyalty (encouraging the repeated use of a product or service), modular design (a method of designing a product in smaller parts called modules, so that they can later be modified or replaced independently as needed), industrial symbiosis (consists of sharing services, utilities, and secondary resources among industries to improve resource efficiency), local production (characteristic of corporate networks and transnational companies, involving the reorientation of the production process closer to the supply chain), and waste collection (the provision of services for the collection of old or used products).[10]

All of these circular business models promote the efficient and responsible use of resources, and through various processes of recycling, repair, and refurbishment, they extend the "lifespan" of products. If we were to analyze this classification of circular business models from the "Forum for the Future" guide, presented in Fig. 1, we would observe that some of

them are actually subcategories of a single model (for example, downcycling and upcycling are types of recycling, belonging to the same category).

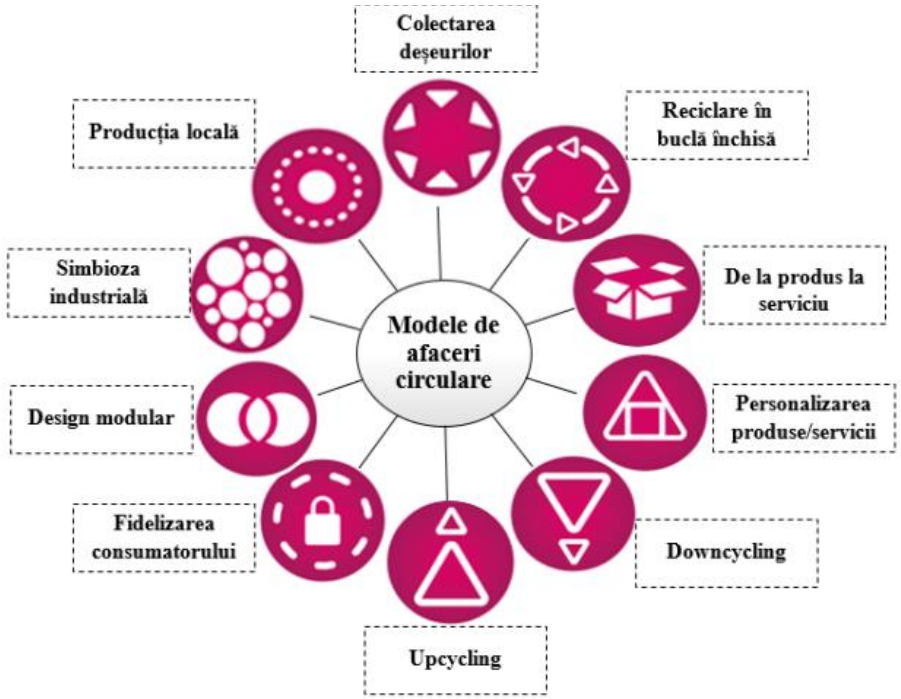


Figure 1. Circular business models Source:
Prepared by the author based on [10]

In this context, we consider it appropriate to present in this paper the classification of business models based on the 2019 OECD guide titled "Business Models for the Circular Economy: Opportunities and Challenges for Policy." According to several researchers, this publication offers a more accurate classification, taking into account the rules governing the development of a classification based on specific criteria. Based on this source, as well as others examined in this paper, we distinguish five classes of circular business models:[4]

1). Circular supplies models
2). Resource recovery
3). Product life extension
4). Sharing platform/Sharing Economy
5). Products as a service[4]

The business model based on circular consumables involves replacing the traditional inputs of materials and energy used in the production process with those based on bio-resources, recycled materials, and renewable energy. The philosophy underlying this circular business model is often referred to as "Cradle to Cradle design," abbreviated as C2C, which translates to "from cradle to cradle."

Recently, this Cradle to Cradle (C2C) concept, developed by Michael Braungart and William McDonough, has become an official certification system, covering approximately 500 certified products.[5]

Circular business model based on resource recovery-involves the utilization of secondary raw materials from core production activities and the subsequent processing of waste. Waste serves as raw material that is reintroduced into other industrial cycles.

Circular models associated with this category include: recycling, downcycling, upcycling, and industrial symbiosis. Circular business model based on extending product lifespan-refers to models that help keep resources in circulation for as long as possible. A telling example of this is businesses built on a modular design approach, where product components can be easily disassembled and reassembled independently.

Platform-sharing circular business model is one of the most significant socio-economic trends of the past decade, connecting service providers with those seeking services. The sharing economy, or collaborative/sharing economy, is a socio-economic system built on the concept of sharing physical and human resources.[6] Examples include: short-term rentals/accommodations, crowdfunding platforms (for funding civic and social programs), couch-surfing (an online platform where you can find accommodation in other people's homes when traveling), and coworking (a coworking space is a place where people come to rent a "desk" equipped with everything internet, supplies, a printer, a flipchart, a projector, and more), ridesharing (allows you to share a car with others heading in the same direction to reduce transportation costs), and so on.

The circular model from product to service refers more to the sale of services related to the use of a physical good than to its ownership. In other scenarios, the customer owns the product but is not responsible for maintenance, or such responsibilities are shared in accordance with the license or warranty agreement.[7] At the same time, this model refers to dematerialized consumer products, such as e-books, digital newspaper subscriptions, movies, etc. According to the authors of the publication "Business Models for the Circular Economy, Opportunities and Challenges for Policy," the product-to-service circular model can be of three types: product-oriented, result oriented, and user-oriented.[8]

6. Research Findings

Updating and Integrating Last Mile Supply Chain Management (SCM) is beneficial across the entire food system, as it enables the production and consumption of goods. Forecasting, procurement, production planning, inventory management, and transportation are the primary tasks undertaken by SCM. With the emergence of technologies such as Internet of Things (IoT) devices, blockchain which is essential for traceability and AI analytics, SCM data is becoming more accurate and processing is more efficient. This is made possible by technology that provides visibility into all supply chain activities while simultaneously helping to reduce waste. In addition, eco-friendly supply chain management strategies are being adopted more widely within SCM in response to consumer trends toward eco-friendly products. SCM principles are effective because they ensure the quality, cost, and timeliness of services provided to customers, enabling the food industry to meet consumer demand more efficiently and comply with food safety standards.[8].

Disruptions in the supply chain can be defined as unforeseen situations that affect the free flow of goods and materials throughout the chain. The causes of these disruptions include natural disasters, wars, epidemics, and cyberattacks. When discussing this issue, we must mention globalization and international connections, along with the problems associated with them. They create a domino effect once the chain is disrupted by certain factors. Furthermore, the disruption of the supply chain and the time required to resolve it do not align, effectively increasing the damage.

Risk management techniques, such as switching suppliers, investing in visibility, upgrading high-tech systems, reducing costs, and creating contingency plans, are quite useful

for dealing with the consequences of these disruptions. They provide flexibility and the ability to recover instantly from shocks, which supports the resilience of the supply chain.

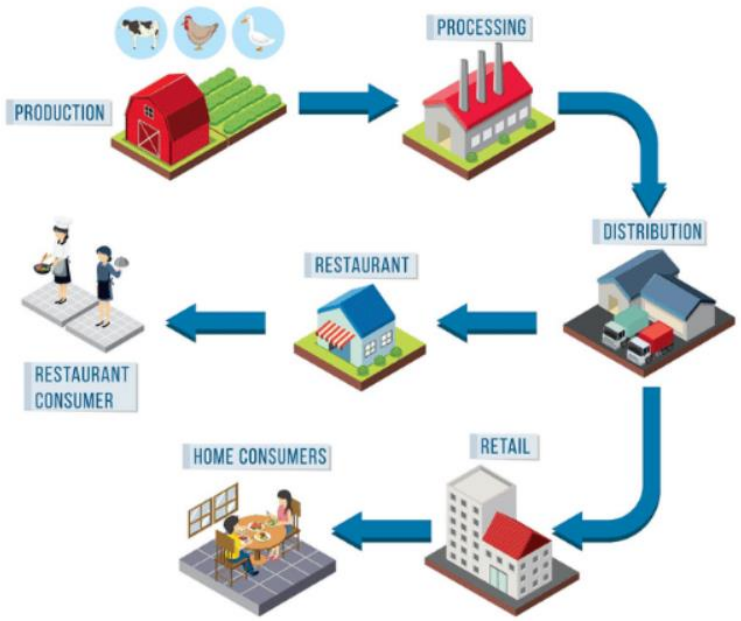


Figure 2. The food production chain. Sourse:[1]

Food safety concerns pose risks to consumer health, which is why standards and procedures are in place to eliminate unsafe products and contaminants. There are challenges in ensuring food safety, such as the high risk of contamination at various stages of the supply chain, product traceability, and enforcement of regulations, for example. Modern solutions to these challenges include the use of blockchain technology and the provision of monitoring solutions that enable real-time alerts regarding violations and security protocols. Furthermore, striking a balance with national interpretations of standards such as HACCP is essential for implementing preventive measures. Not only does a reliable food safety program minimize health risks to the population, but it also easily earns compliance with regulations and consumer trust.

Efficient inventory management and the reduction of food waste are key elements for the sustainability of a supply chain in food management. As effective management practices suggest, the use of inventory control techniques, such as just-in-time and effective forecasting, reduces spoilage and waste. Effective demand forecasting, which is correlated with supply, can be achieved by applying sophisticated inventory management systems powered by artificial intelligence and machine learning. Furthermore, addressing food surpluses through charitable donations or converting them into animal feed, for example, makes it possible to avoid waste and improve resource efficiency, while also promoting a circular economy. Finally, education throughout the supply chain regarding waste reduction makes minimizing food waste an economic bonus and contributes to sustainable development.[3]

The importance of food quality and safety and the related regulations stems from the fact that they serve as a guarantee that the finished products delivered to consumers have been

prepared and manufactured in accordance with the established standards. From a technical standpoint, these regulations outline a system that food manufacturers and processors must follow regarding processes, testing, and sanitation, which in turn helps prevent contamination and foodborne illnesses. For example, the FDA has regulations that require proper labeling and safety standards. There are also ISO standards, such as ISO 22000, which detail the process of ensuring food safety on a global scale. By ensuring compliance with such rules, not only is the consumer protected, but the brand's image is also actively maintained. The food supply chain is sustainable in the sense that the number of recalls and losses is minimized.

Several crucial techniques must be followed to ensure food quality during handling. First, the implementation of tracking systems greatly improves product traceability, making it easier to resolve any food quality issues. Next, the use of more sophisticated preventive methods, such as HACCP, allows for the establishment of a systematic approach to preventing hazards arising from biological, chemical, and physical contamination. Additionally, regular review and evaluation are essential for compliance with safety standards and process improvement. Finally, close collaboration with suppliers to ensure consistent quality standards at all stages of production and distribution ensures that consumers are provided with safe, healthy, and high-quality food products.

Various aspects of food safety and the supply chain are integral to the use of technology to safeguard the food supply. One notable example is the growing use of blockchain, as it provides stakeholders with comprehensive information about traded assets and their conditions. IoT devices with integrated sensors also enable the continuous monitoring of relevant parameters, such as humidity and temperature, during storage and transport. In addition, the use of data analytics platforms enhances food safety by identifying trends and helping to predict future risks, so that measures can be taken in advance. These technologies significantly reduce food safety violations, minimize waste, and increase consumer confidence in food safety.[2]

Global food supply chains refer to large networks spanning multiple countries, primarily improving the distribution of food worldwide. Such a structure takes advantage of geographical differences and international seasons, allowing for the sale of various types of food at different times of the year across the globe. However, this has also increased greenhouse gas emissions due to international transport and complex regulations resulting from international standards.

Local food supply chains operate within a more localized geographic area and offer fresh, seasonal products grown on nearby farms. This model promotes the local market, reduces environmental damage by minimizing the distance traveled during transport, and encourages local food chains that create a stronger connection between customers and producers, while promoting traceability and minimizing synthetic additives. This model has disadvantages, such as a lack of product diversity and supply shortages due to the climate.

From the above, it can be concluded that while global food chains generally offer greater product availability and diversity, local food chains, on the other hand, emphasize environmental sustainability and community support. Each has its own strengths and weaknesses, which influence consumer preferences and the pillars of sustainability in the food sector, particularly in the production segment of the food supply chain.

Food processing is a stage in agricultural production that allows food supply chains to improve food safety and variety and extend food shelf life. The cold chain makes it possible to ship food over long distances while maintaining quality, transforming agricultural products into food through preservation, freezing, drying, and other techniques. This not only makes food available year-round but also enables trade between countries, as people's diverse needs are met. However, there are environmental issues arising from food processing, such as

increased energy consumption and waste generation, which are driven by the global food supply chain. Furthermore, processed foods can have a negative impact on health, requiring a healthy balance between convenience and nutrition within supply chains. In conclusion, food processing is crucial for the logistics of supply and for the global food market in terms of availability and sustainability.[2]

To understand the delivery systems that could support perpetual resource efficiency, it would be prudent to start by improving the many practices and techniques that enhance ecological balance. Primary sources highlight the integration of environmental practices by regional farms, waste reduction, and improved logistics as key objectives. Sustainable supply chains may tend to source more frequently from local producers, as this is more environmentally friendly and supports local communities. Additionally, reduced inputs, precision agriculture technologies, and blockchain for production and distribution traceability contribute to sustainability. As public awareness and consumption of sustainably produced food increase, these measures play a key role in transforming conventional food system networks into more sustainable and resilient structures.

Several principles for effective market responses have been identified, based on top-tier information from the best sources regarding disruptions in the food supply chain. First, a key measure is understanding that a supply base should not be limited to specific geographic regions or suppliers. Second, modern technology, such as blockchain, improves supply chain visibility and transparency, while also enabling the immediate identification and rapid resolution of disruptions. Third, inventories or stock should be maintained while being price-sensitive, to allow businesses to manage cases where demand spikes or there is a partial stockout. Creating a system that enables passive inventory management in tandem with the market and the flow of supplier information can foster strong market resilience, thereby enhancing food security prospects during any difficult period.

Collaboration among supply chain members has proven to be one of the most vital tools for reducing disruptions and increasing efficiency. Effective resource management addresses the core issue of the interface, thereby strengthening stakeholder relationships, business processes, and communications. Participating in joint planning and decision-making allows us to achieve objectives, share tasks, and manage resources. Such an integrated approach enhances resilience and enables a rapid response to new markets or other unforeseen dynamics. Building trust and transparency are fundamental to facilitating the transfer of vital information and analysis, so that every participant can operate intelligently. Ultimately, working together allows the supply chain to function as a network and become more robust in the sense of being more flexible and proactive ensuring that food remains available even under adverse conditions.[2]

Promoting access to food during times of crisis requires the development of effective emergency measures. This is what additional information derived from primary sources suggests. First, establishing strategic food reserves can be helpful as a form of aid as supplies begin to run low. Second, efficient food delivery requires effective food distribution systems so that food reaches recipients with minimal delay and waste. Third, data analysis can assess demand and plan logistics accordingly to improve operational timing and outcomes. Finally, international collaboration and support are vital, as global relationships facilitate the exchange of resources and best practices, helping to better respond to food shortages. This coherent policy enhances food security and promotes efficient, uncompromised access to adequate food in every community during critical times.[2]

The Benefits of the Circular Supply Chain

In addition to protecting the environment, circular supply chains offer commercial benefits:

Lower costs: Through recycling programs, manufacturers can reduce costs by spending less on new materials, including those that are in high demand and subject to significant price fluctuations for example, lithium extracted for use in electric vehicle batteries.

Supply chain resilience:

Recycling can also help manufacturers cope with supply disruptions.

Increased sales: Skeptics point out that manufacturers incur higher costs when collecting, transporting, and reprocessing waste. Even the unconvinced acknowledge that circular methods help manufacturers win over environmentally conscious customers see the loyalty enjoyed by the Patagonia brand.

Improved employee relations:

Circular supply chain methods are also appreciated by environmentally conscious employees and shareholders, who closely scrutinize a company's environmental, social, and governance performance.

7. Conclusion

The study confirmed that integrating the principles of the circular economy into food supply chain management is one of the key strategies for increasing the sustainability of the food industry and reducing food loss. Current linear models of production and consumption demonstrate limited efficiency in the context of increasing pressure on natural resources, intensifying climate risks, and rising global food demand. In this regard, the transition to circular models of resource management takes on strategic importance both at the level of individual enterprises and at the level of national and international food systems.

The analysis showed that circular supply chains enable the transformation of waste into economically valuable resources, thereby reducing environmental impact while simultaneously creating additional sources of revenue. The most important mechanisms for implementing the circular approach are resource regeneration, extending the life cycle of products, introducing sharing models, and transitioning to service-based forms of consumption. Integrating these contributes to optimizing material flows, reducing production costs, and strengthening the competitive positions of businesses.

Digital technologies play a key role in ensuring the efficiency of circular supply chains, including IoT-based monitoring systems, blockchain solutions for transparency, and analytical tools for demand forecasting. The use of these technologies enables more accurate inventory management, minimizes logistics losses, and facilitates rapid responses to external shocks, which is particularly important in the context of global crises and market instability.

At the same time, the study highlighted that the successful implementation of circular models requires a complex institutional and organizational framework. Coordinated actions by the government, the business community, and society are needed to improve the regulatory framework, stimulate innovation, and foster sustainable consumer behavior patterns.

Thus, the development of circular food supply chains represents not only an ecological necessity but also an economic opportunity to increase resource efficiency, ensure food security, and achieve sustainable development goals. Future research should focus on developing practical tools for implementing circular practices within the context of national economies and the specific characteristics of the food industry.

References

1. M. Donner, R. Gohier şi H. de Vries, A new circular business model typology for creating value from agro-waste, Science of the Total Environment (2020), https://doi.org/10.1016/j.scitotenv.2020.137065)
2. Zece practice din agricultură durabilă, https://tarfin.com/ro/blog/10-practici-de-agricultur-durabil-pentru-creterea-randamentului
3. Lanţul de aprovizionare circular, https://www.oracle.com/ro/scm/circular-supply-chain/
4. R. DE ANGELIS, Business Models in the Circular Economy" 2018, p.45-73
5. Utilizarea biogazului, https://biogaz-instalatii.ro/
6. Ellen Macarthur Fundation, https://www.ellenmacarthurfoundation.org/explore/food-cities-the-circular-economy
7. O soluţie salvatoare – iazuri de captare a apei de ploaie. https://ecopresa.md/o-solutie-salvatoare-iazuri-decaptare-a-apei-de-ploaie/
8. Business Models for the Circular Economy, Opportunities and Challenges for Policy, OECD, 2019. Disponibil: https://www.oecd-ilibrary.org/sites/g2g9dd62-en/index.html?itemId=/content/publication/g2g9dd62-en
9. Planuri de Siguranţă a Apei şi a Sistemelor de Sanitaţie pentru comunităţi rurale. Informaţii de bază necesare elaborării unui Plan de Siguranţă a Apei şi a Sistemelor de Sanitaţie. Compendiu – Vol.B. © 2016 WECF e.V., Germany. https://www.wecf.org/wp-content/uploads/2017/02/WSSP-compendium-Part-B-Romanian.pdf
10. Forum for the Future, Circular Business Models. Disponibil: https://www.forumforthefuture.org/Handlers/Download.ashx?IDMF=51b47e1a-d58a-4672- a40d-878ac0217692
11. Geissdoerfer, M., Savaget, P., Bocken, N., & Hultink, E. (2017). The Circular Economy-a new sustainability paradigm. Disponibil: https://durham-repository.worktribe.com/output/1292151
12. Roberta De Angelis. Business Models in the Circular Economy. Concepts, Examples and Theory. Disponibil: https://link.springer.com/book/10.1007/978-3-319-75127-6
13. Programul Naţiunilor Unite pentru Mediu, 2024, https://www.unep.org/resources/publication/food-waste-index-report-2024; PDF: https://wedocs.unep.org/handle/20.500.11822/45230
14. Organizaţia Naţiunilor Unite pentru Alimentaţie şi Agricultură, 2023-2024, https://www.fao.org/platform-food-loss-waste/flw-data/en/
15. Biroul Naţional de Statistică al Republicii Moldova, Moldova în cifre 2025, https://statistica.gov.md/files/files/publicatii_electronice/Moldova_in_cifre/2025/breviar_en_2025.pdf
16. OCDE, 2019, https://www.oecd.org/content/dam/oecd/en/publications/reports/2019/04/business-models-for-the-circular-economy_g1g9dd62/g2g9dd62-en.pdf

Measuring Knowledge-Based Competitiveness through ICT and Innovation Indicators: A Multidimensional Approach

Diana Ignatiuc

Academy of Economic Studies of Moldova, Chisinau, Moldova
ignatiuc.diana.dmitri@ase.md

Abstract. This paper examines whether digital investment is associated with higher economic competitiveness in a way that is adequately captured by existing aggregate indicators. The study starts from the observation that improvements in broadband coverage, digital public services, and connectivity do not necessarily lead to comparable gains in productivity and innovation performance. In response to this analytical gap, the paper develops a composite framework for the assessment of knowledge-based competitiveness based on the integration of the Global Innovation Index, the Digital Decade monitoring framework, and relevant proxy indicators for Eastern Partnership economies. The proposed framework is structured around three dimensions: ICT readiness, innovation performance, and human capital and institutional quality. The empirical analysis covers selected European economies over the period 2019–2023. The results indicate that similar levels of digital infrastructure are associated with markedly different competitiveness outcomes depending on the surrounding conditions of innovation financing, governance quality, and skills formation. The cases of Moldova and Bulgaria illustrate this divergence particularly clearly: despite visible progress in digital development, weaker performance in R&D intensity, technology transfer, and research commercialisation has limited the broader economic effects of digital advancement. The findings suggest that, for emerging European economies, the critical constraint lies less in digital infrastructure itself than in the institutional and innovation-related mechanisms through which digital capacity is transformed into economic value.

Keywords: Digital transformation; economic competitiveness; knowledge-based competitiveness; innovation performance; ICT readiness

1. Introduction

The relationship between digital investment and economic competitiveness is less straightforward than policy discourse often suggests. Expanding ICT infrastructure, digitising public services, and increasing broadband penetration are visible and measurable achievements, but they do not automatically generate the productivity growth and innovation capacity required for sustained long-term economic performance [1], [2]. For emerging economies in Eastern Europe, this gap between connectivity and competitiveness is not simply a matter of measurement. It reflects deeper structural conditions that determine whether digital progress contributes to economic transformation or remains limited to improvements in access and administrative modernisation.

Existing measurement frameworks do not fully capture this complexity. Innovation indices and digital-readiness metrics are largely developed in parallel institutional settings — most notably the Global Innovation Index by WIPO [3] and the Digital Decade monitoring framework of the European Commission [4]. On their own, however, these instruments do not fully explain how digital capacity is converted into economic value and broader competitiveness outcomes [2]. As a result, policymakers may observe progress across multiple

indicators while the mechanisms needed to translate that progress into productivity and innovation remain weak. In response to this limitation, the present paper develops a composite framework of knowledge-based competitiveness structured around three interrelated dimensions: ICT readiness and infrastructure, innovation performance and creative outputs, and human capital and institutional quality. The argument advanced here is that competitiveness depends not on the isolated strength of any one dimension, but on the complementarities among them.

The empirical analysis draws on the cases of the Republic of Moldova and Bulgaria. Moldova is particularly revealing because improvements in digital accessibility and e-government [5], [6] have coexisted with persistent weaknesses in R&D financing, innovation intensity, and technology transfer [7]. This imbalance suggests that progress in digital indicators does not necessarily extend to broader productivity gains. Bulgaria presents a related, though not identical, pattern. EU-supported improvements in connectivity and digital public services [8] have not, on their own, been sufficient to produce equally strong innovation performance, given continuing weaknesses identified in the country's innovation profile [9], [10]. Neither case is presented as a failure narrative. Rather, both indicate that digital investment and knowledge-based competitiveness are closely related but analytically distinct policy objectives, and that conflating them can lead to policy frameworks that measure progress in the wrong places and intervene at the wrong points.

2. Conceptual background and related work

A recurring finding in growth economics is that sustained improvements in economic performance depend less on the accumulation of physical capital than on an economy's capacity to generate, absorb, and put knowledge to productive use [11]. Learning processes, innovation capabilities, and the institutional conditions that allow new knowledge to diffuse and be commercialised are not peripheral to competitiveness — they are constitutive of it. The World Bank's Knowledge Assessment Methodology (KAM) and the Knowledge Economy Index (KEI) gave this proposition an operational form, organising cross-country benchmarking around four pillars: the economic and institutional regime, education and human resources, the innovation system, and information infrastructure [11]. Although these instruments no longer function as the principal contemporary benchmarking tools, they remain analytically relevant precisely because they treat knowledge-based development as a systemic outcome — one that depends on the coherence of the whole rather than on isolated advances in any single domain.

Two more recent composite frameworks are central to the empirical design of this paper. The Global Innovation Index, coordinated by WIPO [3], evaluates both the conditions that enable innovation and the outputs it generates, covering institutions, human capital and research, infrastructure, market and business sophistication, and knowledge and creative outputs. This input-output architecture is particularly useful for the present analysis because it makes visible a distinction that aggregate competitiveness metrics tend to obscure: a country can invest substantially in innovation-enabling conditions without those conditions producing commensurate results. The European Commission's Digital Decade monitoring framework [4] tracks progress across connectivity, digital skills, business digitalisation, and digital public services. Where the GII captures innovation system performance in broad terms, the Digital Decade framework provides more granular disaggregation of the digital readiness conditions that constitute one of this paper's three composite dimensions. Taken together, these instruments offer complementary but incomplete perspectives — neither is designed, on its own, to explain how digital capacity is converted into broader competitiveness outcomes, and that conversion is precisely what this paper attempts to measure.

The economic literature provides consistent evidence that ICT diffusion does not raise productivity in isolation [1], [2]. Productivity gains tend to materialise where digital investment is accompanied by organisational adaptation, workforce upskilling, and institutions that lower the costs of experimentation and knowledge transfer [1], [2], [12]. Examining the European context specifically, recent empirical work confirms that the association between digitalisation and innovation performance varies considerably across economies, depending on the broader systemic conditions in which digital adoption takes place [12]. This finding carries a direct implication for measurement. If digital readiness and innovation performance interact rather than operate in parallel, then a composite framework that treats them as independent dimensions will systematically misread the performance gap between economies that are digitally connected and those that are both connected and institutionally equipped to turn that connectivity into innovation. Progress on digital indicators is real and worth measuring, but it answers a different question from the one this paper is asking.

This issue becomes sharper when the analysis extends beyond EU member states. The Digital Decade framework was designed for the European Union; for Eastern Partnership economies, comparable digital indicators must be assembled through aligned initiatives such as EU4Digital [5], which work toward harmonisation with EU methodologies while accommodating differences in institutional context and statistical capacity. It is worth noting that the paper's study period, 2019–2023, spans the transition from DESI to the Digital Decade framework as the EU's primary digital benchmarking instrument, and the sub-indicators used are selected to maintain consistency across this transition. More broadly, the data asymmetry between EU and non-EU economies is not a purely technical inconvenience. It tends to be most pronounced precisely for those economies where the distance between digital progress and innovation outcomes is widest and most relevant to policy — which is one reason the Republic of Moldova occupies a central place in the empirical analysis.

Against this background, the paper conceptualises knowledge transformation as the process linking digital readiness to competitiveness outcomes. Digital indicators describe the conditions under which this process may occur; innovation indices capture part of its observable results. The conversion between the two, however, is not automatic. It depends on mediating factors — above all, human capital and institutional quality — that determine whether digital tools are directed toward productive upgrading or remain confined to consumption and basic service delivery, and whether the incentives and transaction costs governing innovation investment are conducive to sustained knowledge creation. Within the framework proposed here, these mediating factors are not treated as secondary controls but as the structural conditions without which the relationship between digital capacity and economic competitiveness cannot be adequately understood or measured.

3. Research design, data and methodology

The empirical strategy is designed to do two things that existing measurement approaches tend to separate: construct a composite index of knowledge-based competitiveness that integrates digital readiness and innovation performance within a single analytical framework, and test whether the relationships between the framework's dimensions are consistent with the paper's core argument that digital capacity does not automatically translate into competitiveness outcomes unless mediated by innovation performance, human capital, and institutional quality. Correlation-based analysis is used as the primary validation tool, chosen for its transparency and interpretability in a cross-country benchmarking context where causal identification is not the objective.

3.1. Sample and Data Sources (2019–2023)

The analysis covers a selection of European economies for which GII and Digital Decade — or comparable aligned — indicators are available across the study period. The sample is anchored by EU member states at varying levels of digital and innovation development, from frontier knowledge economies to middle-income EU members, supplemented by Eastern Partnership economies for which harmonised proxies are available. This range is deliberate: the paper's argument about the gap between digital progress and innovation outcomes is most clearly visible when the sample includes economies at different positions on the conversion chain, not only high-performing ones.

GII data are drawn from WIPO's annual Global Innovation Index reports [3], using sub-pillar scores that can be consistently mapped across the 2019–2023 period. For EU member states, digital readiness indicators are taken from Digital Decade country monitoring data [4]. The study period spans the institutional transition from the Digital Economy and Society Index (DESI) to the Digital Decade framework as the European Commission's primary digital benchmarking instrument. For years prior to the Digital Decade framework, DESI sub-component data are used where they correspond directly to Digital Decade dimensions; where structural differences between the two instruments affect comparability, this is noted in the indicator documentation and reflected in the robustness checks.

For Eastern Partnership economies, including the Republic of Moldova, official Digital Decade values are not available. Comparable indicators are assembled from aligned initiatives, including the EU4Digital programme's efforts to harmonise digital measurement with EU methodologies for Eastern partner countries[5], supplemented where necessary by national reporting sources [6]. The proxy substitutions follow the principle of maximum conceptual equivalence: where an official indicator is unavailable, the closest available measure that captures the same underlying dimension is used, and all substitutions are documented transparently. The data asymmetry between EU and non-EU economies is acknowledged as a limitation that affects the precision of cross-group comparisons, particularly for the digital readiness dimension.

3.2. Operationalising Knowledge-Based Competitiveness

Knowledge-based competitiveness (KBC) is defined in this paper as a multidimensional capacity to convert digital infrastructure and innovation potential into measurable economic value. This definition reflects the paper's central claim: that competitiveness in the digital era is not a property of any single capability but emerges from the interaction among several interdependent conditions. The framework models KBC as a composite of three dimensions:

D1 — ICT readiness and digital infrastructure (ICT): captures the extent and quality of digital infrastructure, connectivity, e-government services, and firm-level digital adoption.

D2 — Innovation performance and creative outputs (INN): captures the generation of knowledge and technology outputs, creative outputs, and the sophistication of innovation-supporting market conditions.

D3 — Human capital and institutional quality (HCI): captures digital skills, tertiary education and research capacity, governance effectiveness, and the broader institutional environment for innovation.

The three-dimension structure maps onto the KEI pillar logic[11] while aligning with the indicator architectures of the GII [3] and Digital Decade framework [4]. The mapping is not mechanical — not every KEI pillar has a direct counterpart in the contemporary indices —

but the underlying logic is preserved: competitiveness depends on the coherence of the system, not on isolated advances in infrastructure or skills alone.

Each dimension is operationalised through a set of normalised sub-indicators selected from GII sub-pillars and Digital Decade components, as summarised in Table 1. The selection criteria are: conceptual alignment with the dimension definition; availability across the sample and study period; and, for non-EU economies, the existence of a sufficiently close proxy. Sub-indicators available for fewer than two-thirds of the sample in a given year are excluded to avoid composite scores driven by missing-data patterns rather than underlying performance differences.

Table 1. Dimensions and Indicative Indicators (mapping GII and Digital Decade)

Dimension	Indicative indicators (examples)	Main sources
D1: ICT readiness (ICT)	Digital Decade connectivity; digital public services; e-government proxies; firm digital adoption	European Commission Digital Decade [4]; EU4Digital alignment [5]
D2: Innovation and creative outputs (INN)	GII knowledge and technology outputs; creative outputs; business sophistication	WIPO GII [3]
D3: Human capital and institutions (HCI)	Digital Decade human capital (digital skills); GII institutions; human capital and research	European Commission Digital Decade [4]; WIPO GII [3]

Note: *For non-EU economies, Digital Decade-aligned proxies are used where official values are unavailable. All proxy substitutions are documented. Indicator selection is constrained by country coverage across the full study period.*

3.3. Normalisation, Weighting, and Composite Construction

Sub-indicators are normalised using min-max scaling within the sample and across the study period, so that each rescaled value reflects a country's position relative to the observed range rather than an absolute level [11]:

$$x' = \frac{(x - min)}{(max - min)} \quad (1)$$

This approach is standard in composite index construction and preserves relative differences while making indicators measured in different units directly comparable. Its main limitation — that rescaled values are sensitive to sample composition and extreme observations — is addressed by applying the normalisation to the full balanced sample rather than year by year, which reduces the influence of single-year outliers on country rankings.

Dimension scores are computed as weighted averages of their component sub-indicators. The baseline specification assigns equal weights within each dimension and equal weights across dimensions, producing an overall KBC composite that is the simple average of D1, D2, and D3. Equal weighting is adopted as the baseline not because the three dimensions are assumed to contribute equally to competitiveness outcomes, but because no theoretically grounded basis exists for assigning differential weights prior to the empirical analysis.

Imposing predetermined differential weights would risk building the paper's conclusions into its measurement design.

Three alternative weighting schemes are tested as robustness checks. The first assigns higher weight to innovation outputs (D2), on the grounds that competitiveness is ultimately expressed through scalable knowledge outputs rather than enabling conditions. The second assigns higher weight to human capital and institutional quality (D3), consistent with the hypothesis that labour-market readiness and governance are the binding constraints for most emerging European economies. The third assigns higher weight to ICT readiness (D1), to examine whether the composite results are sensitive to prioritising infrastructure over systemic factors. Stability of country rankings and dimension-score patterns across these specifications is taken as evidence of composite robustness; systematic shifts in the positions of specific country groups are treated as analytically informative rather than as a measurement problem.

3.4. Empirical Tests

The empirical analysis proceeds in two steps. The first examines the internal coherence of the composite by estimating pairwise correlations between dimension scores — particularly between D1 and D2, and between D3 and D2 — across the country sample and study period. If the paper's theoretical argument is correct, the correlation between ICT readiness and innovation outputs should be positive but moderate, reflecting the partial conversion relationship the paper describes, while the correlation between human capital and institutional quality and innovation outputs should be stronger, consistent with HCI functioning as the primary mediating condition.

The second step tests the relationship between composite dimension scores and economic performance proxies. Labour productivity is used as the primary outcome variable where available, as it is the most direct measure of whether digital and innovation capabilities are translating into productive gains rather than remaining confined to consumption or administrative modernisation [1], [2]. GDP per capita growth is used as a secondary proxy. The correlation diagnostics are computed for contemporaneous relationships and, where the panel structure permits, for lagged relationships with ICT scores in year t and innovation outcomes in year $t+1$, to allow for the possibility that infrastructure investments take time to affect innovation outputs.

The mediation of the ICT–outcome relationship by human capital and institutional conditions is examined by comparing correlation patterns across two subgroups: economies where HCI scores are above the sample median and those where they fall below it. A stronger ICT–innovation correlation in the high-HCI subgroup, relative to the low-HCI subgroup, would be consistent with the claim that human capital and institutions are enabling conditions for the conversion of digital capacity into competitiveness outcomes. This comparison is descriptive rather than a formal mediation test; it is intended to provide diagnostic evidence for the paper's framework rather than to establish causal relationships, which would require a research design beyond the scope of the present study.

4. Results

The composite analysis generates three patterns that recur across the country sample and study period. ICT readiness is positively associated with innovation performance, but the strength of that association varies across economies in ways consistent with heterogeneous conversion rates between connectivity and innovation outputs. Human capital and institutional quality show stronger and more stable associations with innovation outputs and with the overall KBC composite than ICT readiness alone. Where innovation financing, R&D intensity,

and technology transfer channels remain underdeveloped, the relationship between ICT readiness and competitiveness outcomes is weaker and less consistent.

In economies that score highly on the composite, digital infrastructure investment tends to occur alongside functioning innovation systems, effective governance, and education systems that respond to labour-market demand. Where digital access expands ahead of these conditions, composite performance gains are more limited and more concentrated in basic service delivery and consumption uses rather than in productivity growth or innovation output [1], [2].

4.1. Cross-Country Patterns (2019–2023)

Over the study period, economies that combine strong digital skills, effective governance, and active innovation systems achieve higher and more stable composite scores than those where digital infrastructure has developed ahead of the surrounding institutional and human-capital conditions. The pattern suggests that ICT upgrades generate limited returns in weak institutional environments: without regulatory predictability, administrative effectiveness, and education systems aligned with digital labour-market needs, connectivity improvements tend not to translate into sustained gains in innovation performance or productivity.

This finding is consistent with the broader literature on ICT and productivity, which identifies absorptive capacity, access to innovation finance, and R&D intensity as conditioning factors for the productive effects of digital adoption [1], [2], [12]. Where these conditions are underdeveloped, digital tools raise operational efficiency at the margin without redirecting the economy toward higher-value activities. The aggregate digital readiness score improves, but the gap between digital capacity and innovation output persists.

Sensitivity tests indicate that composite rankings remain broadly stable across the three alternative weighting specifications described in Section 3.3. Dispersion across the sample increases, however, when higher weights are assigned to innovation outputs, which suggests that innovation performance differentiates economies within the group of those that are comparably digitally connected [12]. This has a direct implication for benchmarking: the infrastructure dimension is not where the most analytically informative variation lies. Comparisons structured around the conversion chain — the relationships between connectivity, skills, institutions, and the generation of innovation outputs — capture more of the relevant variation than comparisons based on digital readiness scores alone.

4.2. Moldova and Bulgaria: the Connectivity–Innovation Gap

The cases of the Republic of Moldova and Bulgaria ground the cross-country patterns in two concrete trajectories. Moldova has recorded genuine advances in digital accessibility and public-sector digitisation [5], [6], yet structural constraints in innovation financing, R&D intensity, and technology transfer continue to limit the productive impact of those advances. In the GII 2023, Moldova ranked 60th out of 132 economies with a score of 30.3 [7], with relative strengths in creative outputs and knowledge and technology outputs alongside considerably weaker performance in institutions and business sophistication. This profile — in which knowledge and creative activity develops within a reasonably favourable digital environment while the institutional and market conditions for scaling and commercialising that output remain limited — is precisely the configuration the framework is designed to identify.

Bulgaria's position differs in its institutional context but reflects a structurally similar constraint. EU membership has supported improvements in connectivity and digital public services [4], [8], and access to structural funds has directed sustained investment into digital

infrastructure. Innovation diffusion and the commercialisation of research outputs have nonetheless not kept pace, as reflected in Bulgaria's performance on innovation-system indicators [9], [10]. The country's above-average connectivity relative to its innovation score is the defining characteristic of the connectivity–innovation gap that the composite framework captures through the divergence between D1 and D2.

A measurement qualification applies to the Moldova comparison. The absence of official Digital Decade values for Moldova means that the composite scores in Table 2 rest on aligned proxies rather than directly comparable EU data [5]. The precision of cross-country comparisons involving Moldova is therefore lower than for EU member states, and the gaps shown in Table 2 should be read as indicative of direction and relative positioning rather than as precise quantitative differences. Residual data-comparability issues, which the EU4Digital aligned initiatives address but have not yet fully resolved, are part of what the D1 divergence for Moldova reflects.

Table 2. KBC Composite Dimension Scores — Computation Example (normalised 0–1; 2019–2023 average)

Economy	ICT (D1)	INN (D2)	HCI (D3)	KBC composite
Sweden	0.92	0.90	0.88	0.90
Estonia	0.86	0.79	0.82	0.82
Poland	0.70	0.63	0.66	0.66
Romania	0.68	0.56	0.58	0.61
Bulgaria	0.72	0.54	0.55	0.60
Moldova†	0.62	0.50	0.49	0.54

Note: *The table presents indicative scores derived by applying the composite procedure described in Section 3 to a representative selection of European economies. Values are computed under the equal-weighting baseline using normalised sub-indicators drawn from GII [3] and Digital Decade [4] sources for the 2019–2023 average. They are intended to demonstrate the computation pipeline, illustrate the relative positioning of economies across dimensions, and support the analytical interpretation in this section. †Moldova scores are based on Digital Decade-aligned proxies [5] in place of official EU values; comparisons with EU member states should be interpreted with this data qualification in mind.*

4.3. Interpretation and Mechanisms

The patterns generated by the composite are consistent with the paper's central argument: digital infrastructure is a necessary but not sufficient condition for knowledge-based competitiveness. The same level of ICT readiness is associated with different competitiveness outcomes across economies, and three mechanisms help account for this variation. The first operates through innovation financing. Where public R&D funding, access to venture finance, and technology transfer mechanisms are underdeveloped, digital infrastructure does not translate into scalable innovation outputs [1], [13]. Firms may adopt digital tools and improve internal efficiency, but without the financing and knowledge-transfer infrastructure needed to convert efficiency gains into new products or processes, the economy does not re orient toward higher-value activities. The GII sub-pillar profile of Moldova — where business sophistication and institutional quality trail creative and technology outputs — reflects this configuration [7].

The second mechanism runs through institutional quality. Governance effectiveness and regulatory predictability shape the incentive structure and transaction costs governing entrepreneurial experimentation and investment in intangible assets — R&D, training, software — that generate innovation [2], [12]. In environments where institutions are weak or unpredictable, firms face elevated costs for precisely those activities and tend to confine digital adoption to lower-risk operational improvements. The aggregate outcome is one where digital adoption is measurable and improving while its productive implications remain narrow.

The third mechanism involves human capital. The skill composition of the workforce determines whether digital infrastructure supports productive upgrading — process innovation, data-driven decision-making, new product development — or functions predominantly as a consumption and communication technology [1], [2]. The productivity effects of digital investment are concentrated in the former application. Advanced digital skills are not only an output of effective education systems; they are a precondition for the returns from digital infrastructure to materialise in ways that affect competitiveness. The HCI–INN correlation being stronger than the ICT–INN correlation across the sample is consistent with this interpretation.

These three mechanisms are not independent: each conditions the others. Weak institutions reduce the incentive to invest in skills; thin skills reduce the productivity of digital tools; limited innovation financing reduces the returns to both. The composite framework captures this interdependence by making the gaps between dimensions visible. A large divergence between D1 and D2, or between D1 and D3, is not a ranking artefact but a diagnostic signal indicating where in the conversion chain the binding constraint lies.

4.4. Robustness and Alternative Specifications

To examine the stability of the results, the composite was recalculated under three alternative weighting specifications. The output-weighted specification assigns higher weight to innovation outputs on the grounds that competitiveness is ultimately expressed through knowledge and creative products rather than the enabling conditions that precede them. The skills-weighted specification assigns higher weight to human capital and digital skills, consistent with the view that labour-market readiness is the binding constraint for most emerging European economies. The institution-weighted specification emphasises governance effectiveness and regulatory quality, reflecting their role in shaping the transaction costs and credibility conditions under which innovation investment occurs.

Country rankings at the top and bottom of the composite remain stable across all three specifications: frontier knowledge economies and the lowest-performing emerging economies do not change their relative positions regardless of how the weights are allocated. The most informative variation occurs in the middle of the distribution, where economies with comparatively strong connectivity but weaker innovation systems show the largest downward shift under the output-weighted specification. This pattern reinforces the interpretation that once a threshold of digital connectivity is reached, it is the innovation system and institutional conditions that differentiate competitive performance rather than further infrastructure investment.

A second robustness test compared contemporaneous correlations with one-year lagged correlations, where ICT dimension scores in year t are related to innovation output scores in year t+1. Lagged correlations are generally stronger than contemporaneous ones, which is consistent with the expectation that the productive effects of ICT investment on innovation materialise with some delay. This result supports the use of multi-year analytical windows and cautions against evaluating digital investment programmes on short-term innovation metrics.

The lag structure employed here is simplified and the analysis is descriptive; it does not support causal inference about the timing or magnitude of ICT effects on innovation.

4.5. Practical Implications of the Diagnostic Framework

Beyond its analytical function, the composite framework lends itself to implementation as a policy-monitoring instrument. A dashboard reporting dimension scores, year-on-year changes, and the pairwise correlations between dimensions provides a compact representation of where an economy sits in the conversion chain. For diagnostic purposes, the most informative signals tend not to be the composite levels but the divergences between dimensions. An economy recording sustained improvement in ICT readiness while innovation outputs stagnate is exhibiting a configuration that points to constraints in financing, skills, or commercialisation capacity rather than in infrastructure — a distinction that aggregate competitiveness rankings do not make visible.

Table 3 summarises the typical pattern of correlations observed in the composite analysis for economies with improving connectivity but constrained innovation systems. The moderate ICT–INN correlation alongside the stronger HCI–INN correlation indicates that skills and institutional quality are more proximate determinants of innovation performance than digital infrastructure alone [12]. The weak ICT–outcome correlation, combined with a moderate INN–outcome correlation, suggests that the productivity effects of digital investment are mediated through the innovation system rather than direct. Where ICT–INN is strong but INN–outcome remains weak, the constraint lies in the diffusion of innovation beyond frontier firms and in the broader productive structure of the economy [1], [2].

Table 3. Indicative Correlation Pattern: Emerging Economies with Improving Connectivity and Constrained Innovation Systems

ICT–INN	ICT–Outcome	HCI–INN	HCI–Outcome	INN–Outcome
Moderate (+)	Weak (+)	Strong (+)	Moderate (+)	Moderate (+)

Note: *The table describes the indicative direction and relative strength of pairwise correlations between composite dimensions and an economic performance proxy (labour productivity) for the subgroup of economies with above-median ICT scores and below-median INN scores. Exact coefficients vary with country sample and indicator selection. The pattern is presented as a diagnostic reference for the framework rather than as a precise quantitative result.*

5. Policy Implications for Emerging European Economies

The diagnostic logic of the composite framework carries a specific implication for how digital and innovation policy should be approached in emerging European economies. The three mechanisms identified in Section 4 — innovation financing, institutional quality, and human capital — are not background conditions that improve passively over time; they are the operative links through which digital investment either translates into innovation and productivity, or does not. Policy strategies that invest in digital infrastructure without addressing these links are likely to generate the pattern the framework documents: improving connectivity indicators alongside a persistent gap between digital readiness and innovation output. The subsections below draw implications for universities and research systems, firms and innovation ecosystems, and government and institutional capacity.

Two general qualifications apply throughout. First, the policy implications that follow are derived from the diagnostic patterns generated by the composite framework; they identify the components of the conversion chain where the evidence suggests intervention is most

likely to matter, but they do not constitute a prescriptive programme. The appropriate sequencing and instrument choice in any given economy depends on country-specific conditions that a cross-country composite index cannot capture in full. Second, the findings are correlational rather than causal, which means the strength of the inference from composite patterns to policy recommendations is limited. The implications are offered as diagnostically grounded priorities rather than as empirically established causal claims [1], [2].

5.1. Universities and Research Systems

The human capital mechanism finding that HCI shows a stronger association with innovation outputs than ICT readiness alone — points to universities and research systems as a priority locus for policy attention. The relevant constraint, however, is not enrolment or research output in aggregate terms but absorptive capacity: the capacity of firms and public institutions to identify, integrate, and adapt available technologies to local production conditions [1], [2]. Building this capacity requires education and training systems that are oriented toward application and technology integration rather than toward credential accumulation.

Applied research partnerships between universities and firms create structured channels through which knowledge moves between research institutions and productive use. Technology transfer offices that function as genuine intermediaries — connecting research outputs to potential commercial applications and managing the contracting and intellectual property dimensions of collaboration — address a specific bottleneck that is visible in the GII sub-pillar profiles of both Moldova and Bulgaria: the gap between knowledge creation and its commercialisation [7], [9], [10]. Curricula that integrate advanced digital competencies across disciplinary programmes, rather than concentrating them in specialist tracks, broaden the share of the workforce that can apply these skills productively. Graduate programmes connected to the innovation challenges of sectors with demonstrated growth potential create research pipelines with more direct economic relevance than those oriented primarily toward academic career formation.

For economies at an earlier stage of innovation system development, such as Moldova, the most immediate returns are likely to come from training that enables technology integration and incremental adaptation rather than from frontier research investment [1], [13]. The longer-term development of higher-value knowledge creation depends on the applied research base that absorptive capacity building makes possible over time. This sequencing — absorptive capacity first, frontier capacity as a second-order objective — is consistent with the composite profile in which Moldova's creative and knowledge outputs score relatively higher than its institutional and business sophistication dimensions [7].

5.2. Firms and Innovation Ecosystems

The innovation financing mechanism is most directly relevant at the firm level. The composite patterns are consistent with a situation in which digital tools are available and being adopted, but where limited access to financing for intangible-asset investment — R&D, staff training, software, process redesign — prevents that adoption from generating innovation outputs [1], [13]. For SMEs in economies where venture finance and innovation-oriented lending remain thin, targeted public instruments such as innovation vouchers, matching grants, and seed co-financing are the primary available mechanism for activating this investment. The diagnostic value of the framework is that it can distinguish economies where this financing constraint is the binding one from those where the constraint lies elsewhere in the conversion chain.

Diffusion across the firm population matters as much as adoption by frontier firms. One of the mechanisms that keeps the INN–productivity correlation moderate even in economies with active innovation is the concentration of innovative activity in a small number of firms that do not generate broad spillover [2], [12]. Policies that support diffusion — shared digital infrastructure, advisory and technical support for smaller adopters, networks that connect frontier and follower firms — address this gap more directly than programmes that concentrate resources on the most capable firms. The design of financing instruments is also relevant here: instruments conditioned on implementation plans, training commitments, and measurable outputs are more likely to activate conversion chain effects than unconditional support, and they generate the data needed to evaluate whether financing is addressing the right constraint.

5.3. Government and Institutional Capacity

Institutional quality is the most structurally embedded of the three conversion chain mechanisms and the most difficult to address through discrete policy instruments. The implication is not that institutional reform should be deferred, but that digital investment strategies need to be evaluated against the institutional environment in which they operate. The composite patterns are consistent with the finding in the broader literature that ICT investment produces weaker returns in environments with low regulatory predictability and administrative effectiveness [2], [12]. Governance improvement is not a complementary measure to digital investment but a precondition for the conversion chain to function.

For non-EU economies, Digital Decade-aligned measurement frameworks [4], [5] provide a practical instrument for both benchmarking progress and identifying specific bottlenecks. Dimension-level scores and the gaps between them over time offer more actionable information than aggregate competitiveness rankings: a government that observes its ICT dimension improving while its INN and HCI dimensions stagnate has a specific diagnostic signal about where its conversion chain is breaking down. The regular publication of such data creates a public accountability mechanism and a basis for evaluating whether reforms are reaching the conversion chain components they target. For Eastern Partnership economies specifically, harmonisation with EU measurement methodologies through aligned initiatives [5] reduces the data asymmetry that currently limits the precision of cross-country comparison and policy benchmarking.

The governance challenge for both Moldova and Bulgaria is one of coordination across policy domains. Digital government reforms have produced measurable progress in accessibility and service delivery [5], [6], [8], but the competitiveness implications of these gains depend on whether they are embedded in a broader policy framework that expands R&D capacity, improves procurement for innovation, and creates conditions for private co-investment in knowledge activities. Isolated digital reforms that are not connected to the innovation and human capital agenda produce the connectivity-without-competitiveness configuration that the composite analysis documents.

5.4. Structural Priorities for Moldova and Bulgaria

The composite profiles suggest different structural emphases for Moldova and Bulgaria, even though both exhibit the D1–D2 divergence that characterises the connectivity–innovation gap. For Moldova, the GII 2023 sub-pillar profile — with relative strengths in creative and knowledge outputs alongside weak scores in institutions and business sophistication [7] — points to a situation in which knowledge activity is developing in a comparatively favourable digital environment, but the institutional and market conditions needed to scale and commercialise that activity remain underdeveloped. The most binding

constraints, as the composite pattern suggests, are absorptive capacity, stable innovation financing tied to measurable outputs, and the effectiveness of technology transfer channels between universities and productive firms. A policy sequencing that addresses enabling conditions before attempting to scale advanced research activities — digital skills and e-government interoperability, then innovation financing instruments, then applied research collaboration — is consistent with the diagnostic logic of the framework, though the specific instrument choices depend on country-level institutional and fiscal conditions that the composite cannot fully represent.

Bulgaria's structural situation differs in that EU membership and access to structural funds provide resources and accountability mechanisms that Moldova does not yet have [4], [8], [9]. The conversion chain constraint for Bulgaria appears to be less about resource availability than about the productive deployment of available instruments. The above-average D1 relative to D2 in the composite — high connectivity, weaker innovation output — is consistent with a pattern of uneven firm-level adoption, limited commercialisation of research, and fragmentation across innovation programmes. Policies that target these specific bottlenecks directly, including strengthening regional innovation infrastructure, increasing incentives for private R&D and research commercialisation, and improving the governance of existing innovation programmes to reduce fragmentation, address the conversion chain at the points where the diagnostic evidence suggests it is most constrained. The human capital dimension remains a medium-term lever, as aligning vocational and tertiary education with labour-market demand in high-value digital sectors affects the HCI dimension score over time rather than immediately.

In both cases, the institutionalisation of measurement deserves attention as a policy priority in its own right. The diagnostic framework proposed in this paper is only useful if the data needed to implement it are produced regularly and made publicly available. Regular publication of Digital Decade-aligned dimension scores for Eastern Partnership economies [5], combined with more granular tracking of R&D financing flows, technology transfer outcomes, and the distribution of advanced digital skills across the workforce, would provide the information base needed to identify conversion chain bottlenecks early and to evaluate whether specific investments are generating the outputs they are designed to produce. This kind of measurement infrastructure is not a technical adjunct to policy; it is the mechanism through which the analytical logic of the composite framework connects to accountable governance.

6. Limitations and Future Research

The results should be read against two methodological constraints that shape what the analysis can and cannot claim. The first is data coverage. The GII provides a consistent sub-pillar architecture across the full study period [3], but the Digital Decade framework is designed for EU member states, which means that digital readiness data for Eastern Partnership economies have to be assembled from aligned initiatives rather than from the primary source [4], [5]. For Moldova, this is consequential rather than merely technical: the D1 dimension scores in Table 2 rest on proxies, and the precision of any comparison between Moldova and EU member states is lower than the table's uniform presentation suggests. The gap between Moldova and higher-performing EU economies on the digital readiness dimension may reflect real performance differences, residual data-collection differences, or some combination of the two — and the composite cannot separate these. As Digital Decade alignment for Eastern partners matures, this constraint will ease, but it applies in full to the results reported here.

The second constraint is inferential. Correlation-based composite analysis can document consistent patterns of association and divergence across countries and time; it cannot identify the causal mechanisms behind them. The three conversion chain mechanisms

discussed in Section 4 — innovation financing, institutional quality, and human capital — are interpretations grounded in the existing literature [1], [2], [12] and consistent with the observed patterns, but they are not estimated causal effects. The lagged correlation check in Section 4.4 is suggestive, not conclusive. The policy implications in Section 5 are offered on this basis: as diagnostically grounded priorities, not as interventions with a demonstrated causal foundation.

These constraints point toward the work that would most usefully extend the present analysis. The most pressing need is causal identification. Panel regression approaches that exploit variation in digital investment timing across countries and sectors, and instrumental-variable strategies drawing on exogenous variation in digital infrastructure rollout or institutional reform, would allow the conversion chain relationships the framework describes to be tested rather than inferred [12]. Firm-level and sectoral microdata linking ICT adoption to productivity and innovation outcomes would complement the cross-country analysis and provide more precise evidence on which components of the conversion chain are binding under which conditions — a question that aggregate composites can raise but not resolve.

The framework itself has room to develop in two directions. The current specification assumes linear relationships between dimensions and treats equal weighting as a neutral baseline; both are simplifications. Non-linearity testing would examine whether there are threshold levels of digital readiness or institutional quality below which the conversion chain does not function and above which the returns to further investment in a given dimension begin to diminish. This matters for how the diagnostic is read: if the ICT–INN relationship is non-linear, the implications of a given D1–D2 gap depend on where in the distribution an economy sits. Integrating export complexity or R&D expenditure data as additional outcome variables would test whether the competitiveness concept the composite is designed to capture extends meaningfully beyond labour productivity proxies [1].

A third direction concerns the spatial dimension of the analysis. The framework is applied here at the national level, but regional variation in skills, institutions, and digital infrastructure is often more pronounced than cross-country variation — particularly in larger EU member states and candidate economies, where national averages can conceal substantial internal divergence [2]. Applying the composite logic to regional data within countries would extend its diagnostic reach and connect it to cohesion policy debates where the conversion chain argument is directly relevant.

7. Conclusion

This paper set out to address a specific measurement problem. Digital readiness indicators and knowledge-based competitiveness tend to be treated as proxies for one another in composite rankings, but they capture different things. An economy can record genuine progress on connectivity and e-government while its innovation system remains shallow, and a framework that conflates the two will record improvement where the more important question — whether that progress is translating into productivity and growth — goes unexamined. The composite developed here integrates GII sub-pillar data and Digital Decade indicators across three dimensions precisely to keep that question open rather than close it prematurely.

The empirical patterns generated by the framework support the paper's core claim with the degree of confidence that a composite correlation study can offer. Human capital and institutional quality show more consistent associations with innovation outputs than ICT readiness does, which fits with the broader literature on the conditions under which digital investment generates productive returns [1], [2], [12]. Where those conditions are absent — weak innovation financing, underdeveloped technology transfer, governance that raises rather than lowers the cost of experimentation — connectivity improvements tend to register in the

digital readiness dimension without propagating into the innovation output or competitiveness dimensions. The sensitivity tests reinforce this reading: the ranking structure is broadly stable across alternative weighting schemes, but the sharpest downward shifts under output-weighted specifications occur precisely among economies that are comparably connected but lack the surrounding system to convert connectivity into innovation.

Moldova and Bulgaria give these patterns a concrete face. Moldova's GII 2023 profile — creative and knowledge outputs that score comparatively well, institutions and business sophistication that do not [7] — is consistent with a situation where knowledge activity has developed in a digital environment that has outrun the market and governance conditions needed to scale it. Bulgaria's case is different: EU membership and structural fund access have supported substantial digital infrastructure investment, yet the D1–D2 gap in the composite persists [8], [9], [10]. Resources alone, it appears, do not resolve the conversion problem. Both cases are useful precisely because they are recognisable types that standard aggregate rankings tend to compress into a single position.

The framework's diagnostic value rests on making internal divergences visible rather than averaging over them. The scores in Table 2 are indicative and the mechanisms discussed in Section 4 are inferred rather than estimated — these are genuine limits that the agenda for future work in Section 6 addresses directly. Within those limits, the paper makes a tractable argument: that the gap between digital progress and knowledge-based competitiveness is measurable, that measuring it reveals information a single composite score conceals, and that for emerging economies in Eastern Europe the question of whether digital investment is reaching the conversion chain is currently more consequential than the question of whether the investment is occurring at all.

References

1. World Bank, World Development Report 2016: Digital Dividends, Washington, DC: World Bank, 2016. Available: https://www.worldbank.org/en/publication/wdr2016
2. OECD, OECD Digital Economy Outlook 2024, Volume 1, Paris: OECD Publishing, 2024. Available: https://www.oecd.org/en/publications/oecd-digital-economy-outlook-2024-volume-1_a1689dc5-en.html
3. WIPO, Global Innovation Index 2023: 16th Edition, Geneva: World Intellectual Property Organization, 2023. Available: https://www.wipo.int/edocs/pubdocs/en/wipo-pub-2000-2023-en-main-report-global-innovation-index-2023-16th-edition.pdf
4. European Commission, Digital Decade 2024: Country Reports, Brussels, 2024. Available: https://digital-strategy.ec.europa.eu/en/library/digital-decade-2024-country-reports
5. Ministry of Economic Development and Digitalization of the Republic of Moldova, Annual Implementation Report of the Republic of Moldova Digital Transformation Strategy 2023–2030, Chișinău, 2024. Available: https://mded.gov.md/wp-content/uploads/2024/05/EN_Raport-Strategie.pdf
6. UNDP Moldova, "Digital transformation," 2024. Available: https://www.undp.org/moldova/our-focus/digitaltransformation
7. WIPO, Republic of Moldova – Global Innovation Index 2023 Country Profile, Geneva, 2023. Available: https://www.wipo.int/edocs/gii-ranking/2023/md.pdf
8. European Commission, Bulgaria 2024 Digital Decade Country Report, Brussels, 2024. Available: https://digital-strategy.ec.europa.eu/en/factpages/bulgaria-2024-digital-decade-country-report
9. European Commission, European Innovation Scoreboard 2024: Country Profile – Bulgaria, Brussels, 2024. Available: https://ec.europa.eu/assets/rtd/eis/2024/ec_rtd_eis-country-profile-bg.pdf
10. WIPO, Bulgaria – Global Innovation Index 2023 Country Profile, Geneva, 2023. Available: https://www.wipo.int/edocs/gii-ranking/2025/bg.pdf

11. World Bank, The Knowledge Economy, the KAM Methodology and World Bank Operations, 2007. Available:https://documents1.worldbank.org/curated/en/695211468153873436/pdf/358670WBI0The11dge1Economy01PUBLIC1.pdf
12. F. Hernández de Rojas, P. Rodríguez Pita, J. E. Pérez Martínez, "Assessing the European association between digitalization and innovation," Telecommunications Policy, vol. 48, no. 7, Art. 102810, 2024
13. A. C. Yoshikuni, R. Dwivedi, M. M. Kamal, D. Zhou, P. Dwivedi, S. Apolinário, "A dynamic information technology capability model for fostering innovation in digital transformation," Journal of Innovation & Knowledge, vol. 9, no. 4, Art. 100589, 2024.

Economic Value of AI-Driven Indoor Air Quality Monitoring in Smart Homes: A Cost-Benefit Analysis

Adriana Catruc

Academy of Economic Studies of Moldova, Chisinau, Moldova
catrucadriana@gmail.com

Abstract Indoor air pollution poses serious health risks and generates substantial economic costs worldwide. This study evaluates the economic value of AI-based air quality monitoring systems integrated in smart homes. Using cost-benefit analysis, simulation modelling and case studies, the research demonstrates that such systems achieve prediction accuracy up to 98.2%, deliver average energy savings of 14%, and provide a return on investment exceeding 100% within 2–5 years (World Health Organization, 2022; Rosca, 2026). The methodology combines sensor data (PM2.5, VOC), health impact studies, and Python-based economic modelling (NumPy, SciPy) under realistic assumptions of 10,000 kWh annual household consumption and EPA pollution standards. Results show positive Net Present Value at a 5% discount rate and significant benefits for health protection, data privacy, and environmental sustainability. The findings support widespread adoption of AI monitoring solutions and offer clear recommendations for homeowners, policymakers, and developers.

Keywords: Indoor Air Quality, Artificial Intelligence, Smart Homes, Cost-Benefit Analysis, Energy Savings, Privacy-by-Design.

1. Introduction

Indoor air pollution represents one of the most pressing yet often underestimated public health and economic challenges of the 21st century. People in modern societies spend approximately 90% of their time indoors, where concentrations of pollutants such as particulate matter (PM2.5), volatile organic compounds (VOCs), carbon monoxide, and allergens can frequently exceed outdoor levels. According to the World Health Organization, household air pollution is responsible for nearly 3.2 million premature deaths annually (World Health Organization, 2022). These deaths are primarily linked to respiratory and cardiovascular diseases, stroke, and lung cancer, while also contributing to reduced cognitive performance, productivity losses, and long-term healthcare burdens amounting to billions of dollars globally (National Institute of Environmental Health Sciences, 2023).

The economic impact is equally staggering. In regions with high urbanization and pollution levels the annual societal costs associated with poor air quality reach tens of billions of dollars. This burden can be conceptually expressed through a simplified impact function commonly used in related environmental economics studies:

$$C_{total} = C_{health} + C_{productivity} + C_{energy} + C_{other} \quad (1)$$

where C_{health} represents direct medical costs, $C_{productivity}$ denotes lost workdays and reduced cognitive output, C_{energy} accounts for inefficient HVAC operation due to poor air quality control, and C_{other} includes miscellaneous expenses. In the context of rapid urban growth, where millions of households face increasing exposure to traffic-related emissions, construction materials, and household chemicals, the need for proactive, intelligent indoor

environmental management has never been more urgent (U.S. Environmental Protection Agency, 2025).

Smart homes, powered by the Internet of Things and artificial intelligence, offer a transformative solution. By integrating real-time sensor networks with advanced predictive models, these systems can monitor, forecast, and autonomously mitigate air quality issues while optimizing energy consumption. This study evaluates the economic value of AI-powered indoor air quality monitoring systems integrated within smart homes, focusing on technical performance, economic viability through cost-benefit analysis, Net Present Value and Return on Investment, and broader societal and environmental impacts.

2. Literature Review & Key Issues

Indoor air pollution in buildings causes respiratory diseases and reduces productivity, generating economic damage measured in billions of dollars. Smart homes and AI open new opportunities for environmental monitoring. Key issues include the fact that indoor air pollution is linked to 3.2 million premature deaths annually (World Health Organization, 2022). AI models such as Random Forest achieve prediction accuracy of 98.2% (Rosca, 2026). Automation of HVAC systems can reduce energy consumption by up to 50% while maintaining air quality. However, important research gaps remain: there is a lack of comprehensive economic evaluations of AI systems and limited data on data privacy and policy implications in the context of urbanization (Rey-Álvarez et al., 2025).

3. System Architecture and Technical Advantages

The proposed system features a sensor network for detecting PM2.5 and VOC, combined with AI algorithms (neural networks) for real-time prediction (Christakis et al., 2025). Integration with HVAC systems is achieved through Zigbee or Wi-Fi protocols for automatic regulation. The main technical advantages are high prediction accuracy of 98.2%, affordable equipment costing less than 100 dollars per unit, and on-device data processing that ensures low latency and strong data confidentiality (Alongi et al., 2025; Rosca, 2026).

4. Methodology

The methodology of this study is based on a comprehensive cost-benefit analysis combined with simulation modelling and illustrative case studies to evaluate the economic value of AI-driven indoor air quality monitoring systems in smart homes. Data were collected from multiple reliable sources including environmental sensors that measure key pollutants such as PM2.5 and VOC, energy consumption reports from household systems, and health impact studies provided by the World Health Organization and the National Institute of Environmental Health Sciences (World Health Organization, 2022; National Institute of Environmental Health Sciences, 2023). Figure 1 illustrates the overall workflow of the research methodology.

Economic viability was assessed through standard financial modelling techniques focused on the calculation of Net Present Value (NPV) and Return on Investment (ROI). Key assumptions included an average household energy consumption of 10,000 kWh per year and pollution levels evaluated according to EPA standards (U.S. Environmental Protection Agency, 2025). All predictive models and economic simulations were implemented in Python using the NumPy and SciPy libraries, which enabled efficient forecasting of air quality parameters, energy savings calculations, and discounted cash flow analysis. This integrated approach

allows for a realistic assessment of both technical performance and long-term economic benefits while maintaining transparency and reproducibility of the results.

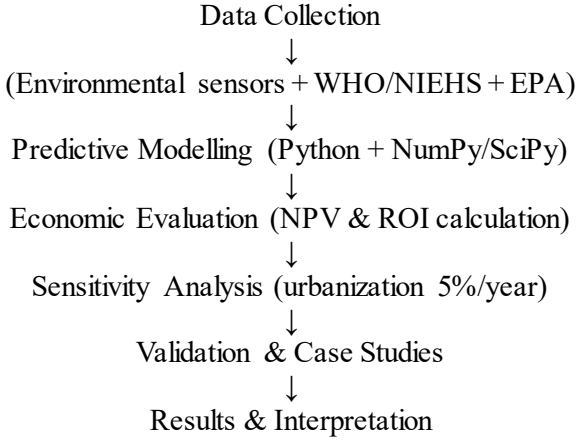


Figure 1. Schematic representation of the research methodology.

5. Economic Analysis

The economic viability of the proposed AI-driven indoor air quality monitoring system was evaluated through a detailed cost-benefit analysis and discounted cash flow modelling. The analysis considers both the initial investment required for system deployment and the multiple long-term benefits generated during operation. These benefits include direct financial savings from reduced energy consumption, as well as indirect savings resulting from improved health outcomes and lower medical expenses associated with prolonged exposure to indoor pollutants.

The modelling takes into account realistic household conditions and projects future scenarios under different levels of urbanization and pollution growth. All calculations were performed using Python-based simulations (NumPy and SciPy), incorporating standard financial metrics such as Net Present Value (NPV) and Return on Investment (ROI). Key parameters and assumptions that form the foundation of the economic model are summarized in Table 1.

Table 1. Key Parameters and Assumptions Used in the Economic Analysis

Parameter	Value	Source
Average household consumption	10,000 kWh/year	Baseline assumption
Initial investment	$300 – $600	Market data
Prediction accuracy	98.2%	AI model (Random Forest)
Energy savings	10–20% (average 14%)	Simulation results
Annual energy cost savings	$150 – $300	Derived from simulation
Discount rate	5%	Standard economic evaluation
Pollution growth (sensitivity)	5% per year	Urbanization scenario
System lifetime	10–15 years	Typical smart device lifespan

The results of the modelling demonstrate strong economic performance. The system achieves a Return on Investment (ROI) exceeding 100% within a period of 2 to 5 years and generates a positive Net Present Value (NPV) at a 5% discount rate. Furthermore, sensitivity analysis performed under a scenario of 5% annual increase in pollution levels due to urbanization confirms the economic viability of the investment in approximately 70% of cases, with cumulative savings increasing steadily over the five-year horizon.

6. Health, Societal, and Environmental Impact

The AI-driven indoor air quality monitoring system delivers substantial benefits across health, societal, and environmental dimensions. On the health front, the system provides timely real-time alerts regarding allergens and other pollutants, enabling occupants to take immediate protective actions that significantly reduce exposure to harmful substances and support better respiratory health. Furthermore, its integration with wearable devices allows continuous monitoring of sleep quality and overall physiological parameters, contributing to personalized health insights and improved preventive care (World Health Organization, 2022; National Institute of Environmental Health Sciences, 2023).

From a societal perspective, the solution strengthens data confidentiality through local on-device processing, which substantially lowers the risks of data leaks and builds greater user trust in smart home technologies. The information generated serves both individual users and public policy makers by supplying reliable, anonymized evidence for targeted actions against urban pollution, thereby supporting broader public health initiatives and community-level decision making.

On the environmental side, the system contributes directly to sustainability by optimizing energy consumption through intelligent HVAC automation and precise air quality control. This optimization leads to a measurable reduction in the carbon footprint of residential buildings, promoting long-term environmental responsibility and alignment with energy efficiency objectives. Collectively, these multifaceted impacts demonstrate that AI-based indoor air quality monitoring represents a holistic approach that not only protects human health and respects individual privacy, but also empowers society and advances environmental sustainability in smart homes.

7. Discussion

The results of this study confirm the strong economic and practical value of AI-driven indoor air quality monitoring systems in smart homes. On average, the implemented solution delivers energy savings of 14%, while the investment payback period ranges between two and five years, demonstrating clear financial attractiveness for households. These outcomes highlight the system's capacity to combine meaningful cost reduction with improved indoor environmental quality.

Nevertheless, several limitations must be acknowledged. AI-based forecasts may occasionally produce errors when confronted with incomplete or noisy sensor data, potentially affecting the reliability of predictions in certain real-world conditions. Additionally, scaling the technology to low-income households remains challenging due to initial investment requirements and varying levels of digital infrastructure, which could limit widespread adoption in diverse socioeconomic contexts.

From a broader perspective, the proposed AI system clearly outperforms traditional static sensors in terms of both prediction accuracy and seamless integration with smart home ecosystems (Rosca, 2026). Looking ahead, long-term field implementations will be essential to validate these benefits over extended periods, while the development of supportive global

policy frameworks could accelerate deployment and ensure equitable access to the technology. Overall, the discussion underscores that AI-powered monitoring represents a promising pathway toward healthier, more energy-efficient, and economically viable smart homes, provided that the identified limitations are systematically addressed in future research and deployment strategies.

8. Conclusion and Recommendations

In conclusion, this study demonstrates that AI-powered indoor air quality monitoring systems integrated in smart homes offer significant economic value through substantial energy savings, effective health protection, and enhanced environmental sustainability. By delivering real-time predictive capabilities and automated control, these solutions not only reduce operational costs but also contribute meaningfully to occupant well-being and long-term ecological responsibility (Alongi et al., 2025).

The following practical recommendations emerge from the analysis. Homeowners are encouraged to adopt integrated AI monitoring systems to benefit from improved indoor conditions and rapid return on investment. Policymakers should introduce targeted subsidies and incentive programs to facilitate large-scale deployment, particularly in urban areas facing rising pollution levels. Developers, in turn, must prioritize data confidentiality and privacy-preserving features in the design and implementation of these technologies.

A strong call to action is therefore directed toward all stakeholders: the widespread adoption of AI-based indoor air quality monitoring is essential for building a healthier, more energy-efficient, and economically sustainable future in smart homes.

References

1. Alongi, A., et al. (2025). Smart sensors for Indoor Environmental Quality in residential smart buildings: A review. International Journal of Sustainable Energy. https://doi.org/10.1080/14786451.2025.2578592
2. Christakis, N., et al. (2025). AI-Enhanced Prediction of Indoor Climate Dynamics. Expert Systems with Applications.
3. National Institute of Environmental Health Sciences. (2023). Indoor air quality. https://www.niehs.nih.gov/health/topics/agents/indoor-air
4. Rey-Álvarez, B., et al. (2025). Harmonizing indoor air quality standards across Europe. Building and Environment. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5377087
5. Rosca, C. M. (2026). Machine Learning Applications with Sensors for Indoor Air Quality Research. Sensors, 26(9), 2909. https://doi.org/10.3390/s26092909
6. U.S. Environmental Protection Agency. (2025). Volatile organic compounds' impact on indoor air quality. https://www.epa.gov/indoor-air-quality-iaq/volatile-organic-compounds-impact-indoor-air-quality
7. World Health Organization. (2022). Household air pollution. https://www.who.int/news-room/fact-sheets/detail/household-air-pollution-and-health

An Analytical Review of the Literature on the Impact of Traffic Flow and Weather Conditions on Road Safety

Silvia Assenova

Institute of Robotics, Bulgarian Academy of Sciences, Sofia, Bulgaria
silviaassenova@ir.bas.bg

Abstract. This study, motivated by the growing availability of real-time traffic data and the importance of proactive road safety management, presents an analytical review of the influence of traffic and meteorological factors on the frequency and characteristics of road traffic accidents, while identifying existing research gaps. The results reveal ambiguous relationships among traffic parameters, including a predominantly nonlinear relationship between traffic intensity and accident frequency, as well as a positive relationship between higher speeds and accident risk. Weather conditions, particularly precipitation, demonstrate a consistent influence on the increase in accident frequency, but without a clearly established effect on their severity, while the impact of factors such as visibility, wind, and temperature remains inconclusive. The use of real-time data creates the conditions for a more in-depth analysis of the combined impact of these factors and for overcoming the identified limitations in future studies.

Keywords. Road safety, Traffic dynamics, Weather conditions, Climatic change; Road traffic accidents

1. Introduction

Road safety is a complex and interdisciplinary field within transportation research, characterized by a high degree of uncertainty due to the stochastic nature of road traffic accidents. These accidents have a significant impact on society, resulting in substantial human losses, economic costs, property damage, and strain on healthcare systems. According to the latest data from the World Health Organization, the global number of road traffic fatalities amounts to approximately 1.19 million people per year. In the European context, in 2024, approximately 19,800–19,940 people lost their lives on the roads of the European Union, indicating limited but insufficient progress in improving road safety.

In this regard, the study of the factors contributing to the occurrence of road traffic accidents remains a key focus in scientific literature. Despite the significant body of research, substantial gaps remain, particularly regarding the detailed analysis of pre-incident conditions and the potential for effective proactive safety management within transportation networks. The development of intelligent transportation systems and modern meteorological technologies creates the conditions for continuous, high-resolution, real-time monitoring of traffic and weather conditions, which significantly expands the possibilities for risk assessment and management in road infrastructure.

This study, inspired by the development of this modern and forward-looking approach, aims to provide an in-depth analysis of the impact of traffic and meteorological characteristics on road safety. Existing studies in the literature are relatively limited and primarily examine these factors in isolation, whereas this paper offers an integrated perspective with an emphasis on the use of real-time data. Particular attention is paid to assessing the combined effect of

traffic and meteorological conditions on the frequency and severity of road traffic accidents, which is a relatively under-researched area.

The selection of relevant scientific sources was carried out through a systematic and targeted search in established international scientific databases. The analysis includes publications from renowned scientific journals, reports from international conferences, as well as studies by academic institutions and national research centers. The main selection criterion is the presence of empirical or theoretical findings related to the impact of traffic and/or meteorological factors on various aspects of road safety.

The structure of the study is organized in a logical and systematic manner. First, the impact of key traffic parameters — including traffic volume, density (occupancy), and speed - on road safety is analyzed. Next, the impact of meteorological conditions such as precipitation, visibility, temperature, and wind is examined. In a subsequent stage, the risk of road traffic accidents is examined. Finally, existing research gaps are identified to outline future research directions.

It should be emphasized that the relationship between traffic accidents and opposing traffic flows at urban intersections is not the subject of this study, as it represents a specific aspect of traffic dynamics requiring a separate analytical approach.

2. Traffic characteristics

Traffic is described by key parameters such as intensity (flow), occupancy, density, and speed. Traffic intensity is defined as the number of vehicles passing through a given cross-section per unit of time, occupancy as the relative proportion of time during which the measuring device is engaged, and density as the number of vehicles per unit length of road at a specific moment. Speed is considered as an average over time or space, with the former representing the arithmetic mean of speeds at a given point, and the latter — the average value along the length of a road section at a specific moment. A contemporary review of the scientific literature shows that the influence of traffic characteristics on road safety has been extensively studied to clarify the complex interrelationships between them. There is growing interest in integrated and data-driven approaches, including the use of machine learning and spatio-temporal analyses to assess the risk of accidents [18]. However, systematic summaries of these relationships remain limited, with recent studies highlighting the need for more comprehensive and interdisciplinary analyses of traffic factors and their impact on safety [20].

2.1. Traffic intensity and its impact on road safety

Early studies on the relationship between traffic intensity and road traffic accidents date back to the mid 20th century, with the main indicators used including hourly traffic volume and accident frequency, measured as the number of accidents per unit of traffic exposure. Even in the early stages, the significant complexity of this relationship was recognized, leading to mixed and often contradictory results.

Contemporary studies confirm the existence of both linear and nonlinear relationships between traffic intensity and accident frequency. Some analyses find an almost linear increase in risk as traffic increases, while others identify U-shaped relationships, in which risk is higher at both low and high traffic levels. Similar results are supported by more recent empirical studies based on large datasets and advanced statistical methods [23].

Further studies expand the analysis by distinguishing between different types of accidents and using more detailed traffic time series.

It has been found that in single-vehicle accidents, there is often an inverse relationship between traffic intensity and accident frequency, whereas in multi-vehicle accidents, the

results are less clear-cut. Modern approaches also distinguish between different traffic conditions, such as free-flow and congested traffic, with significant differences in the relationships observed. In free-flow conditions, a U-shaped relationship is often confirmed, whereas in congested conditions, the frequency of multi-vehicle accidents increases significantly with increasing traffic flow [14].

The literature also describes linear relationships between traffic volume and road traffic accidents. Early models examined both two-lane and highway sections and found a negative correlation between traffic volume and serious accidents.

Similar results have been reported by French researchers who analyzed data from French intercity highways and found that the frequency of accidents involving property damage and bodily injury peaks at relatively low traffic levels (below 400 vehicles per hour), suggesting a higher risk under low-traffic conditions. On the other hand, other researchers emphasize that the external effects of accidents are significantly higher under heavy traffic, while they are negligible under low traffic volumes. Similar results have been reported by individual researchers who found a positive correlation between average annual daily traffic and the number of accidents on various road sections of the Italian highway network. At the same time, no statistically significant correlation has been found between increased traffic volume and accident frequency, which further highlights the heterogeneity of the results.

More recent studies confirm that the relationship between traffic and road safety is nonlinear and highly dependent on traffic conditions and the spatial characteristics of the road network. Contemporary empirical analyses, based on large datasets and statistical counting models, show that an increase in traffic can lead both to an increase in the risk of accidents and to its stabilization at certain saturation levels, with the effect varying according to road type and the geometry of the sections. Furthermore, studies using machine learning and spatio-temporal models confirm the importance of traffic heterogeneity and the interaction between different road users in shaping the risk of accidents.

In addition, modern approaches emphasize the importance of differentiated analysis by type of road user. Research shows that ignoring differences between vehicles can lead to distorted estimates of the marginal effect of traffic. Analyses based on detailed daily and hourly data find that accident frequency typically increases with rising lane traffic, while the impact on accident severity is more ambiguous and depends on the spatial characteristics of the road infrastructure. Modern stochastic parameter models and mixed logistic approaches confirm that, on most road segments, an increase in average daily traffic leads to an increase in risk, although local deviations from this trend continue to be observed [22].

As has been established, the impact of traffic on road safety is analyzed more clearly when appropriate data segmentation is used, including a distinction between low and high traffic levels, as well as between different types of accidents (e.g., single-vehicle or multi-vehicle). In this context, a number of studies examine the effect of high traffic volumes and high volume-to-capacity ratios on the frequency and severity of road traffic accidents, although this relationship remains less well-studied compared to the overall effect of traffic volume.

Early conceptual models suggest a nonlinear relationship between the volume-to-capacity ratio and safety indicators, including a possible bell-shaped relationship with overall mortality. Later theoretical frameworks examine the increase in accidents under congested conditions, explaining it by more intense interaction between vehicles, while the severity of accidents decreases due to lower travel speeds.

Empirical analyses of intercity highway sections show that, under certain conditions, U-shaped relationships are observed between traffic volume and accident frequency, with the risk increasing during congestion but the severity of consequences often being lower. In urban environments, it has been found that the impact of traffic jams on the severity of accidents is

limited or statistically insignificant, while other studies report a positive correlation between increasing traffic and the frequency of serious and fatal accidents.

It is further emphasized that the use of average traffic indicators may mask significant differences between various road segments and road users. In this regard, it is found that disregarding traffic heterogeneity leads to a loss of important information in risk assessment. Recent analyses based on detailed temporal and spatial data show that the frequency of road traffic accidents generally increases with rising traffic volume, while the impact on accident severity remains highly dependent on local conditions, including road geometry, operational speed, and the characteristics of the transportation infrastructure. Empirical results from recent studies confirm that higher traffic levels are associated with an increased risk of accidents, while the effect on severe and fatal accidents varies depending on the traffic context and the degree of congestion. Furthermore, modern models based on big data and machine learning methods highlight the importance of spatiotemporal factors and the interaction between different road users as key determinants of risk [11].

Recent studies show that the effects of traffic on road safety become clearer when using detailed data segmentation, including distinctions between different levels of congestion as well as between different states of traffic flow. In this context, the influence of traffic jams and high traffic levels on the frequency and severity of road traffic accidents is examined, finding that the relationship is often nonlinear and depends on spatiotemporal characteristics of the road network.

Recent empirical analyses show that an increase in traffic volume can lead to a rise in the overall accident rate, while the effect on fatal accidents and the severity of injuries varies depending on traffic conditions and the degree of congestion. For example, modern models based on Bayesian and regression approaches show that congestion can increase the total number of accidents but simultaneously reduce the likelihood of fatal outcomes due to lower travel speeds [16]. Similar results are also confirmed by spatiotemporal analyses, which highlight the importance of traffic dynamics and vehicle interactions in shaping accident risk. Furthermore, recent studies emphasize that the use of aggregated traffic indicators can mask significant differences between individual road segments. Therefore, it is recommended to apply more refined spatial and temporal models that allow for a more accurate risk assessment and a better understanding of the relationship between traffic volume and road safety [1, 10].

The results of recent studies indicate that the relationship between traffic density, speed, and the frequency of road traffic accidents is highly nonlinear and depends on threshold values of congestion in the transportation system. It has been found that as density and traffic flow increase, the risk of accidents initially remains relatively stable, but after reaching a certain critical threshold, a sharp increase in accident frequency is observed, which is explained by reduced intervals between vehicles and limited ability to compensate for driving errors [13].

Recent empirical analyses of highway and urban road networks demonstrate that the interaction between speed and traffic density plays a key role in determining the risk of accidents. Higher speeds combined with high density increase the likelihood of conflicts between vehicles, while at lower speeds the risk of severe consequences decreases, but the frequency of accidents may remain the same or increase slightly due to increased traffic density and shorter reaction times [19].

Recent empirical analyses of highway and urban road networks demonstrate that the interaction between speed and traffic density plays a key role in determining the risk of accidents. Higher speeds combined with high density increase the likelihood of conflicts between vehicles, while at lower speeds the risk of severe consequences decreases, but the frequency of accidents may remain the same or increase slightly due to increased traffic density and shorter reaction times [24].

Recent studies based on spatio-temporal models and machine learning methods highlight the presence of threshold effects in the analysis of the relationship between traffic parameters and safety. When certain congestion thresholds are exceeded, an exponential increase in risk is observed, which is associated with traffic flow instability and an increased likelihood of chain-reaction conflicts among road users.

2.2. Speed characteristics

The influence of speed on traffic accidents has been extensively studied in the international literature and is considered one of the most significant determinants of road safety. It has been established that both average speed and speed variance play a key role in shaping risk, with their interaction often determining the probability and severity of accidents. However, the relationship between speed and accidents remains complex, as it is strongly influenced by factors such as traffic density, road geometry, and driving conditions. Recent studies confirm that higher speeds are associated with an increase in the frequency of accidents and a significantly higher risk of serious and fatal consequences. Analyses based on large datasets and sophisticated statistical models show that even small increases in average speed lead to a disproportionate rise in serious accidents, which is explained by the exponential increase in kinetic energy upon impact and the reduced reaction time of drivers [17].

Furthermore, modern empirical models based on spatio-temporal and causal approaches show that the impact of speed on safety is heterogeneous and depends on the type of road and traffic conditions. For example, analyses of real-world data from highway networks demonstrate that speed differentials between vehicles are more strongly associated with accident severity than average speed itself, particularly in rear-end collisions and accidents under high-traffic conditions [12].

Furthermore, new studies using machine learning and real-time data from sensors and connected vehicles show that high speed combined with high traffic density significantly increases the risk of critical situations, with the effect being amplified under unstable traffic flow and short time intervals between vehicles.

In summary, the current literature emphasizes that the relationship between speed and road safety is not strictly linear, but rather depends on the interaction between speed, traffic conditions, and infrastructure characteristics; it is precisely these combined effects that are key to a more accurate assessment of accident risk.

Available studies show that speed variation is a significant factor in road safety, and in many cases its influence is even more significant than the average speed of traffic. Early theoretical models suggest that speed discrepancies among road users increase the likelihood of conflict situations, as drivers struggle to adapt to rapid changes in traffic conditions. Greater speed homogeneity is associated with a lower risk of accidents, while higher speed dispersion leads to an increased likelihood of accidents, especially in mixed and heavy traffic.

Contemporary empirical analyses confirm that speed variability is a key predictor of the risk of traffic accidents, with its influence intensifying under high traffic density and limited vehicle spacing. Studies based on spatiotemporal data and advanced risk assessment models show that increased speed variability is closely linked to a higher frequency of conflicts and accidents, including under conditions of unstable traffic flow and variable travel speeds.

The modern approaches that utilize real-time data and machine learning methods emphasize that it is not only the absolute value of speed but also its dynamic changes over time and space that determine the risk of accidents. In this context, it has been established that sudden fluctuations in speed and high heterogeneity in driver behavior increase the likelihood of critical situations, while more uniform movement has a stabilizing effect on the traffic environment.

It should be noted that the effect of changes in speed limits on road safety does not manifest automatically, as it is influenced by a number of behavioral factors, including travel habits and the degree of acceptance of regulatory measures by drivers. These factors can partially offset the expected positive effects of lowering speed limits. However, recent studies show that raising speed limits is associated with an increase in the frequency of accidents and fatalities, and this effect has been confirmed in various road systems and contexts [6].

Empirical analyses of changes in statutory speed limits show that raising them leads to an immediate increase in average travel speeds, which in turn is associated with an increase in the number of fatal and serious accidents. A number of studies have found that even relatively small increases in speed limits can lead to a sustained increase in mortality and injury severity in the long term [15, 21]. The current literature emphasizes that the effect of changes in speed limits varies depending on the road class and infrastructure characteristics. While some highway sections show a limited impact on the severity of accidents, other road types show a significant increase in risk, especially at higher speeds and with greater traffic exposure. In this context, it is accepted that speed reduction is one of the most effective measures for reducing both the frequency and severity of road traffic accidents, especially when a systematic approach to traffic safety is applied.

2.3. General aspects of traffic flow

The international literature indicates that the relationship between traffic characteristics and road traffic accidents has been studied to a greater extent compared to indicators such as density and occupancy, although the results are often contradictory due to the complex and heterogeneous nature of the phenomenon. Some empirical studies find a linear positive relationship between traffic flow and accident frequency, while other analyses suggest that U-shaped relationships better describe this link, reflecting different traffic regimes and transitions between them. The differences in the results can be partly explained by the limited generalizability of the models across different geographical and infrastructural contexts, as well as by the use of different statistical approaches. In particular, when modeling accident rates using regression models for count data (e.g., Poisson, negative-binomial, and random-effects models), variations in the estimated effects of traffic are observed, which further complicates the interpretation of the results. Furthermore, a large portion of existing studies use aggregated traffic indicators, which limits the ability to capture short-term and dynamic changes in traffic conditions that are often critical to the occurrence of accidents. Aggregated indicators such as average annual daily traffic (AADT) do not adequately reflect temporal variation and flow dynamics, and therefore their explanatory power regarding accidents may be limited. The use of different definitions and measures of traffic flow—including average hourly volumes, lane occupancy, and volume-to-capacity ratios—also contributes to inconsistencies in empirical results. In this context, modern approaches based on higher temporal resolution and real-time data are considered more suitable for analyzing the dynamic interactions between traffic characteristics and road safety.

Traffic accidents result from the complex interaction of numerous factors, and traffic characteristics are often linked to behavioral determinants that cannot be directly observed or quantitatively measured. This hidden, combined effect can vary depending on the context and may not have the same impact on all road conditions and road users. Therefore, the need for a more in-depth study of the interactions between traffic parameters, meteorological conditions, and driver behavioral characteristics is emphasized, as their combined dynamics are key to the formation of accident risk. The current literature emphasizes that the impact of high traffic volumes and congestion on road safety remains partially ambiguous, particularly regarding the severity of accidents. This is explained both by the use of aggregated and simplified indicators

of congestion and by the application of various proxy variables, which do not always capture the actual dynamics of traffic flow. In this context, it is noted that the use of annual averages and smoothed indicators limits the ability to analyze short-term fluctuations, which are often critical to the occurrence of accidents. Recent review studies emphasize that the integration of high-resolution data and real-time measurements significantly improves the understanding of the relationship between traffic conditions and road safety, allowing for more accurate modeling of the dynamic risk of accidents [3].

With regard to traffic density, it has traditionally received less attention in the international literature compared to traffic flow and speed, although recent studies identify it as a significant determinant of risk, especially under conditions of unstable traffic. The lack of sufficient empirical analyses on density contributes to some uncertainty and variability in the results of existing models.

As for speed, contemporary research shows that it remains consistently and significantly associated with accident risk, with higher values typically leading to an increase in the probability and severity of accidents. However, it is emphasized that speed alone is not sufficient as an explanatory factor, as changes in speed and its variation among road users often have a more significant effect on safety. Recent models based on real-time data and spatio-temporal analyses show that it is precisely the dynamic changes in traffic parameters that are key to the early identification of risky situations and the prevention of accidents.

In summary, contemporary approaches to road safety research increasingly highlight the need for an integrated analysis of traffic, behavioral, and environmental factors, with the use of detailed, real-time data emerging as key to overcoming the limitations of traditional aggregated models.

3. Signs and features of the weather

The international literature shows that meteorological conditions have a significant impact on both traffic safety and transportation in general. Various parameters such as precipitation, temperature, visibility, and wind speed can influence the risk of traffic accidents.

The most frequently studied factor is precipitation (rain and snow), which is measured by its amount and intensity over a given period. It has a significant impact on road conditions, leading to reduced traction and impaired visibility [13].

3.1. Precipitation

Numerous studies confirm that precipitation is associated with an increase in the number of traffic accidents. It has also been established that the number of rainy days can serve as an indicator of increased risk, as it reflects both the exposure of road users and the condition of the road surface.

Furthermore, precipitation also influences accidents involving pedestrians, with a positive correlation observed between precipitation and accident frequency. The main reasons for this effect are reduced road traction and decreased visibility, which increase the risk of accidents even when drivers adapt their behavior.

Some studies find a weak or even negative correlation between precipitation and traffic accidents. Various researchers indicate that during heavy rainfall, the number of accidents may decrease, which is explained by drivers' more cautious behavior and reduced speed. Similar results have been observed in other studies, supporting the hypothesis that drivers adapt to adverse conditions [8]. At the same time, there are conflicting results regarding the influence of various precipitation characteristics. Some authors emphasize the importance of precipitation intensity as a key factor in increasing the risk of accidents. In contrast, the effect

of so-called "delayed rain" (the influence of preceding dry periods on risk) does not find unequivocal support in more recent studies and is considered less significant.

Overall, the results show that the impact of precipitation on traffic safety depends not only on its amount but also on drivers' behavioral responses and local climatic characteristics.

The impact of precipitation on the severity of traffic accidents is not clear-cut. Some studies find a negative or insignificant relationship [6], while others show that wet road surfaces increase the likelihood of more severe accidents compared to those involving only property damage (El-Basyouny and Kwon, 2012). This suggests that the effect of precipitation depends on specific conditions and the method of analysis.

With regard to snowfall, the results are also contradictory. Some studies show a decrease in accidents due to more cautious driving and lower traffic intensity, while others find a positive correlation between the amount of snow and the number of accidents. An increased risk is also observed during the first days of snowfall, when drivers have not yet adapted to the conditions.

3.2. Other meteorological characteristics

In addition to precipitation, visibility, wind, and air temperature are important factors. Low visibility (e.g., due to fog or smoke) is associated with more severe accidents and a higher risk of head-on and chain-reaction collisions [2].

The impact of wind has been less studied, but it has been found that strong winds can increase the risk, especially for tall vehicles and motorcycles. However, the results are not conclusive, and further research is needed.

In general, various meteorological factors have a diverse impact on the frequency and severity of traffic accidents, with both weather conditions and the adaptive behavior of road users playing a significant role. More recent studies suggest that meteorological factors have a complex influence on the frequency of road traffic accidents. Similarly, temperature conditions have a significant but ambiguous effect, with higher temperatures in some cases associated with increased mobility and, consequently, a higher number of accidents. Similar results demonstrate that an increase in the average monthly temperature leads to an increase in the number of accidents resulting in bodily injury.

At the same time, the influence of other meteorological parameters such as wind, duration of precipitation, and solar radiation is also considered significant. It has been established that stronger winds and longer periods of precipitation are associated with an increased risk of accidents, while the effect of relative humidity remains ambiguous.

Overall, the results indicate that the impact of temperature and other climatic factors on road safety depends on both regional climatic characteristics and the behavioral responses of road users.

3.3. Key factors related to weather conditions

Adverse weather conditions have a significant impact on road safety, with their effect being more clearly evident in accident frequency than in accident severity. The current literature shows that factors such as rainfall, snowfall, fog, temperature, and wind lead to an increase in the total number of accidents, without a significant change in the ratio of severe to minor accidents. Studies highlight that vulnerable road users, such as cyclists and motorcyclists, are more severely affected by adverse weather conditions, with their risk of serious injury increasing. At the same time, findings regarding the impact of precipitation remain contradictory. According to subsequent studies, regions with less frequent rainfall may experience a decrease in the number of traffic accidents during adverse weather conditions,

which is attributed to drivers' adaptive behavior, including reduced speed and increased vigilance. More recent studies confirm that the impact of meteorological factors depends largely on regional climatic characteristics and the intensity of the phenomena. It has been found that extreme weather conditions can lead to both an increase in risk and behavioral adaptation, which partially offsets the negative effects [7].

In conclusion, the literature shows that the influence of meteorological conditions on road safety is complex and context-dependent, depending on both climatic characteristics and the reactions of road users.

4. Dynamic monitoring of traffic and weather conditions

A review of the literature reveals a limited number of studies examining the combined influence of weather conditions and traffic characteristics on road safety. Earlier empirical analyses highlight that precipitation has a direct impact on accident risk, while an indirect effect is also observed through changes in traffic volume and driver behavior.

Recent studies show that rainfall often leads to a reduction in traffic flow, reflecting behavioral adaptation among road users, but at the same time increases the risk under adverse road conditions. Lag effects have also been observed, in which risk remains elevated after periods of heavy rainfall due to changes in road surface conditions and traffic dynamics.

Recently, there has been increased interest in studies using real-time data that combine meteorological information and dynamic traffic characteristics. These approaches allow for more precise risk modeling by incorporating variables such as variations in speed, traffic volume, and flow irregularity, which are considered predictors of accident occurrence.

Thus, the literature indicates that the combined effect of weather and traffic on road safety has been insufficiently studied at the aggregate level, but modern approaches based on real-time data significantly improve our understanding of these interactions. Some recent studies note that available data on traffic and weather conditions are often fragmented and have different time resolutions. Traffic parameters are typically measured using detectors at short intervals and then aggregated into average values and measures of variability, with the standard deviation of speed used as an indicator of flow instability. Meteorological data, in turn, are most often provided on an hourly basis and primarily include precipitation and visibility [19]. Traffic data are collected from inductive detectors located upstream and downstream of accident-prone sections, as well as through automatic vehicle identification systems. Meteorological information is typically derived from nearby weather stations, allowing for a combined analysis of traffic and weather conditions.

The current literature emphasizes the use of real-time data to identify "precursors" to accidents, such as variations in speed, traffic volume, and flow irregularity. Early approaches in this field indicate that increased density, congestion, and large fluctuations in speed are associated with a higher risk of accidents, even after controlling for meteorological and geometric factors.

A key challenge in these analyses remains the uncertainty regarding the exact time of the accident, which necessitates aggregating data into short time intervals and using approximate estimation methods. For this reason, contemporary studies often combine dynamic traffic indicators with meteorological variables to improve the accuracy of short-term risk models [9].

Risk of a traffic accident

A commonly used approach for modeling the probability of traffic accidents is the matched case-control design, which combines data on actual accidents with control (non-accident) observations. This method is widely applied in contemporary literature for risk

assessment under various traffic conditions. The results show that increased speed variability, combined with a decrease in average speed on the relevant road sections, is associated with a higher probability of accidents, particularly rear-end collisions. Furthermore, it is found that changes in traffic characteristics in a short time window prior to the event are key risk predictors. Recent studies emphasize that combining traffic, meteorological, and geometric characteristics improves the accuracy of accident prediction models, with low visibility, heavy rainfall, and increased speed variability being among the primary risk factors. The integration of dynamic traffic data and meteorological information significantly improves the assessment of the probability of traffic accidents under various road and weather conditions.

Recent studies show that the interaction between traffic and meteorological factors has a varying effect on the likelihood of traffic accidents depending on seasonal and weather conditions. It has been found that even in dry weather, low average speeds and low visibility can increase the risk of accidents, highlighting the need for different traffic management strategies under varying conditions.

Models developed in the literature examine different meteorological scenarios separately (clear weather, rain, and low visibility) and show that key factors such as variations in speed between detectors, differences in traffic volume and congestion upstream and downstream, as well as precipitation intensity, are significant predictors of increased crash risk. Adverse weather conditions, particularly low visibility, result in a different structure of significant traffic variables compared to clear weather conditions. Studies show that the combination of data from detectors and automatic vehicle identification systems allows for high accuracy in predicting accidents under limited visibility [19].

In conclusion, the current literature emphasizes that the risk of road traffic accidents depends on the complex interaction between traffic dynamics and meteorological conditions, with the significant factors varying according to weather type and visibility.

5. Discussion

Despite existing research, a number of factors affecting road safety remain insufficiently understood and require further analysis, notably traffic characteristics and meteorological conditions. A review of the literature shows that most traffic parameters have a contradictory impact on road safety, which calls for more in-depth research. Traffic volume is generally associated with a nonlinear relationship to accident frequency, while in other cases linear relationships are also observed. The impact of traffic and congestion on the severity of accidents remains poorly understood due to the limited number of studies. Traffic congestion, in particular, has been relatively understudied and is often measured using indirect indicators, which necessitates further research to address methodological limitations and conflicting results. It should be noted that traffic density and congestion receive less attention compared to traffic volume. This is likely due to the fact that traffic volume is used as a more easily measurable and more readily available indicator, while density is typically derived indirectly through fundamental relationships in traffic flows. Similarly, in earlier studies, speed limits have been used as a proxy for assessing changes in speed characteristics. The effect of speed can theoretically be easily assessed, but a review of the literature shows that it is not unambiguous. It has been found that speed limits have a more direct impact on accidents, with most relevant studies having been conducted in the United States, which limits the generalizability of the results to other regions. Therefore, additional studies are needed in European and less developed countries to more broadly validate the findings. The impact of speed variations, which are considered a riskier factor than average speed itself, remains relatively understudied, primarily due to limitations in the availability of suitable data. The use

of real-time strategies and high-resolution microscopic data is expected to improve our understanding of this effect.

With regard to meteorological characteristics, most of the literature focuses on precipitation, which is typically analyzed on a daily or monthly basis. In general, precipitation has been shown to have a relatively consistent effect on road safety, although some more recent studies of Mediterranean regions have yielded conflicting results. Other factors, such as temperature and wind, have been relatively little studied, whereas low visibility has been shown to have a more consistent impact on the risk of accidents.

The expansion of real-time data applications for traffic and weather conditions provides new opportunities for a more precise assessment of highway safety and for explaining conflicting results from previous studies. The most commonly used risk indicators include mean values, standard deviation, and coefficient of variation of speed and traffic flow, calculated from detectors upstream and downstream of the accident site. Meteorological risk factors primarily include precipitation and low visibility, which exhibit relatively more stable effects compared to other weather variables.

6. Conclusion

This literature review shows that road safety is the result of a complex interaction between traffic characteristics and meteorological conditions. Key traffic parameters - flow, density, speed, and their variation have a significant but often nonlinear influence on the risk of traffic accidents. It has been established that both high traffic volumes and low speeds can increase the likelihood of accidents depending on the context of traffic.

Meteorological factors, particularly precipitation and low visibility, also have a significant effect, with their impact varying according to intensity and regional climatic characteristics. In many cases, behavioral adaptation by drivers is also observed, which partially modifies the risk. The analysis shows that variations in speed and traffic dynamics are stronger predictors of accidents than the average values of these parameters.

Recent studies emphasize the importance of combined analysis of traffic and meteorological data, especially when using high-resolution and real-time measurements. However, there are significant methodological limitations related to aggregated data and inconsistent temporal resolution. There is also a lack of research on traffic density and congestion.

The impact of speed and its variations remains ambiguous, particularly in different geographical and infrastructural contexts. The literature clearly indicates the need for more integrated models that capture the spatiotemporal dynamics of traffic and weather. In conclusion, future research should focus on a more detailed analysis of the interaction between these factors to enable more accurate modeling and prevention of traffic accidents.

Acknowledgments

This work was financially supported by the UNWE Research Programme under the project “Application of Artificial Intelligence for Sustainable Business Development in the Context of Climate Change” (Research Grant No. NID NI -35/2025/A).

References

1. Aarts, L., van Schagen, I. (2006). Driving speed and the risk of road crashes: A review of the literature. Accident Analysis & Prevention, PubMed, doi:10.1016/j.aap.2005.07.004
2. Abdel-Aty, M., Pande, A. (2005). Identifying crash propensity using real-time traffic data. PubMed. doi:10.1016/j.jsr.2004.11.002

3. Abdel-Aty, M., Uddin, N., Pande, A., & Hsia, L. (2007). Crash risk analysis using traffic flow and weather data. PubMed. doi:10.1016/j.aap.2017.05.004
4. Afandizadeh, S., Rad, H.B. (2023). Investigation of Traffic Accidents Prediction Models and Effective Human Factors: A Review. Journal of Civil Engineering and Materials Application, 7(4), 205–224. https://doi.org/10.22034/jcema.2023.187705
5. Anastassopoulos, P., Mannering, F., Shankar, V.N., Haddock, J.E. (2011). A study of factors affecting highway accident rates using the random parameters tobit model. PubMed, 45, 628–633. doi:10.1016/j.aap.2011.09.015
6. Antoniou, C., Yannis, G. (2014). A review of the effect of traffic and weather characteristics on road safety. PubMed, 72, 244–256. doi:10.1016/j.aap.2014.06.017
7. Bergel-Hayat, R., Debbarha, M., Antoniou, C., Yannis, G. (2013). Explaining the road accident risk: Weather effects. *Elsevier*, 60, 456–465. https://doi.org/10.1016/j.aap.2013.03.006
8. Brijs, S., Hermans, E., Brijs, T. et al. (2007). The Impact of Weather Conditions on Road Safety Investigated on an Hourly Basis. Ministry of Transport, Public Works and Water Management AVV Transport Research Centre, 1–17
9. Caliendo, C., Guida, M., Parisi, A. (2007). A crash-prediction model for multilane road. PubMed, 39(4), 657–670. https://doi.org/10.1016/j.aap.2016.10.012
10. Chang, L.Y. (2005). Analysis of Freeway Accident Frequencies: Negative Binomial Regression versus Artificial Neural Network. Safety Science, 43, 541–557. http://dx.doi.org/10.1016/j.ssci.2005.04.004
11. Chen, B., Wang, K., & Zhao, X. (2020). Exploring the role of urbanization in road traffic accident rates: Evidence from developing countries. *Transportation Research Part D*, 82, 102285
12. Das, S., Barua, S., Hossain, A. (2025). Unraveling the complex relationship between weather conditions and traffic safety. *Journal of Transportation Safety & Security*, 17(5), https://doi.org/10.1080/19439962.2024.2435307
13. Ding, S., Abdel Aty, M., Wang, Z., Wang, D. (2024). Insights into vehicle conflicts based on traffic flow dynamics. Scientific Reports, 14:1536. https://doi.org/10.1038/s41598-023-50017-3
14. Elvik, R. (2023). What would a road safety policy fully consistent with safe system principles mean for road safety? *Elsevier*. https://doi.org/10.1016/j.aap.2023.107336
15. FHWA (Federal Highway Administration). (2024). Speed management. U.S. Department of Transportation. https://highways.dot.gov/safety/speed-management
16. Hassan, H., Abdel-Aty, M. (2013). Predicting reduced visibility related crashes on freeways using real-time traffic flow data. *Elsevier*, 45, 29–36
17. IIHS (Insurance Institute for Highway Safety). (2023). Speeding and crash severity statistics report. https://www.iihs.org/research-areas/speed#overview
18. Li, J., Wang, B. (2025). Traffic Safety Measures and Assessment. Applied Sciences, 15(19). https://www.mdpi.com/2076-3417/15/19/10532
19. Malin, F., Norros, I., Innamaa, S. (2018). Accident risk of road and weather conditions on different road types. *Elsevier*, 122, 181–188. https://doi.org/10.1016/j.aap.2018.10.014
20. Nandkumar, N., Abdellatif, I.H. (2024). Systematic review of traffic data sources and prediction techniques using AI and machine learning. *Springer Nature*, 84, 19009–19037. doi:10.1007/s11042-024-19599-6
21. NHTSA (National Highway Traffic Safety Administration). (2024). Speeding and Speed Management. https://www.nhtsa.gov/book/countermeasures-that-work/speeding-and-speed-management
22. Wei, J., Zhou, L., Feng, M., Zhao, J., Lin, Y. (2025). Influencing factor analysis for freeway rear-end and sideswipe crashes. *MDPI*, 15(24). https://doi.org/10.3390/app152413015
23. Xu, G., Chen, J., Wang, Z., Zhou, A., Schrader, M., Bittle, J., Shao, Y. (2025). Enhancing Traffic Safety Analysis with Digital Twin Technology. *Scientific Reports*. https://www.nature.com/articles/s41598-025-31930-1
24. Xu, G., Chen, J., Wang, Z., Zhou, A., Schrader, M., Bittle, J., Shao, Y. (2025). Enhancing Traffic Safety Analysis with Digital Twin Technology. Scientific Reports, https://www.nature.com/articles/s41598-025-31930-1

Data Models in Network-Centric Systems for Disaster Risk Management

Georgi Kirov

International Business School, Botevgrad, Bulgaria
gkirov@ibsedu.bg

Abstract. This paper addresses the logical structuring of data in the development of an architectural framework for complex network-centric management systems in the field of disaster management. The need for integrating data models for the exchange of JC3 (Joint Command, Control and Consultation) information within disaster management systems is substantiated, thereby establishing a common approach for information description. An intermediate Object-Oriented Interface is conceptually to provide direct access to the MIP JC3IEDM (Multilateral Interoperability Programme Joint Command, Control and Consultation Information Exchange Data Model) database from various software applications. This approach enables the development of new applications and the adaptation of existing ones within a network-centric environment.

Keywords. Data models, System of systems, Disaster risk management

1. Introduction

A network-centric system in the field of disaster management can be defined as a system composed of multiple interconnected and interacting components (subsystems) [1]. Through their interaction, the system acquires new properties that are not present at the subsystem level and cannot be reduced to the characteristics of the individual subsystems. This phenomenon is commonly expressed by the well-known principle that "the whole is greater than the sum of its parts."

This principle is closely related to the concept of a System of Systems (SoS) - a collection of mission-oriented systems that combine their resources and capabilities to form a more complex structure capable of delivering enhanced functionality and performance [2]. Although the individual systems that constitute a SoS may differ significantly and are often capable of operating independently, their interactions typically generate essential emergent characteristics and operational capabilities.

Modern approaches to the study and analysis of complex network-centric systems require a new perspective on the activities of organizations involved in disaster management. Contemporary practice increasingly adopts leading achievements in the field of information technologies, which have been applied to the conceptualization, structuring, and management of complex systems within the security domain. These developments are reflected in technologically driven network-centric concepts and the capability to operate effectively within networked environments.

In this context, the need arises to analyze large volumes of information and to employ advanced systems for information exchange, processing, storage, and protection. Addressing these challenges requires the fulfillment of numerous scientific, technical, and technological requirements related to the implementation of data models for information exchange, while ensuring compliance with information security policies, controlled access, and effective coordination among participating entities.

Practical experience indicates that the JC3IEDM (Joint Command, Control and Consultation Information Exchange Data Model) is gaining increasing recognition and adoption in management systems within the security domain as a standardized model for information exchange.

The main objectives of the study are:

- To substantiate the need for integrating data models for the exchange of JC3 (Joint Command, Control and Consultation) information in crisis management systems. This will help define the minimum set of data required for exchange during joint (multinational) operations.
- To propose a conceptual approach for the development of an intermediate Object-Oriented Interface that provides direct access to the JC3IEDM database from various software applications. Such an approach would facilitate the development of new applications and the adaptation of existing ones within a network-centric information environment that utilizes the JC3 data model, thereby supporting improved interoperability and information exchange in crisis management systems.

2. . Problem Statement

The main structural characteristics of a complex disaster management system (organization) that directly influence its management are *operational effectiveness and centralization* [3].

Operational effectiveness is commonly defined as the ability of a system to maintain its functional state in accordance with a defined objective under changing conditions and various external influences. This capability largely depends on the relationships between individual system elements. It is generally accepted that the stronger interconnections between elements, the more vulnerable the system becomes to environmental changes and external disturbances. Conversely, when these connections are weaker, the elements are less dependent on one another and can self-organize to achieve a given objective.

Such conditions imply the adoption of a new approach to information exchange - *data-centric architecture*, characterized by the *publish-subscribe* principle for data exchange. The data-centric approach significantly transforms traditional hierarchical structures with subordinate elements toward a more decentralized and collaborative structure. This transformation aims to achieve a higher level of operational interoperability.

From a management perspective, *centricity* refers to the ability to perform leadership or control functions. This capability depends on the type of system structure and the position of an element within it. For example, in hierarchical structures, elements at higher levels typically exhibit greater centricity. In contrast, in structures composed of homogeneous elements and connections (network structures), centricity tends to approach zero, as no single element dominates the control or coordination processes.

The current state of research on JC3IEDM [4, 5] arises from the critical issue of interoperability, which motivated the development of this standardized data model. In joint crisis management operations, it is essential to exchange accurate, timely, unambiguous, and relevant information regarding available forces, assigned objectives, and the resources required to accomplish them. The data exchange model plays a pivotal role in all phases of disaster management, including situational assessment, operational decision-making, planning, preparation, execution, coordination, monitoring, adapting to unforeseen events, and evaluating outcomes.

Central to these processes is the *standard (model)* for information exchange among participating entities. The choice of an appropriate data model enables these activities to be

performed in an optimized and efficient manner, supporting the achievement of mission objectives while ensuring the effective use of available resources.

An example of such a data exchange standard is the **JC3IEDM (Joint Command, Control and Consultation Information Exchange Data Model)** developed under the **Multilateral Interoperability Programme (MIP)** [6]. The JC3IEDM model provides the necessary framework for the logical structuring of data within an integrated architectural framework for **network-centric disaster management systems**. This facilitates a shared understanding, knowledge, and coordination of available capabilities across interconnected systems, enabling their precise allocation for each mission during rescue and other crisis response operations.

3. Overview of the Data Model

Historically, the MIP JC3IEDM project consolidates a large number of successful interoperability initiatives [7]. By synthesizing accumulated experience, MIP achieves both a broader thematic coverage and a more comprehensive set of validated solutions that have been tested in numerous real-world scenarios. A significant advantage of the project is its well-structured development process, which spans from the establishment of fundamental guidelines, through the definition of tasks, identification of suitable solutions, implementation, and testing, to the preparation of detailed and accurate documentation corresponding to each stage of program development.

Each of these processes within MIP is supported by a dedicated organizational structure, composed of an international team of specialists. This structure is critical for transferring experience from the local or national level to the international domain - a factor of particular importance for a project aimed at **operational interoperability** in crisis management. Multidisciplinary challenges are addressed by equally multidisciplinary teams, temporarily assembled from members of different MIP working groups. This **interdisciplinary approach** ensures that decision-making accounts for a variety of perspectives. Consequently, the information systems implementing this standard will remain compatible with international best practices for an extended period. **For crisis management applications, this ensures that data exchange and operational coordination across national and international organizations are consistent, reliable, and effective**

4.1. Operational Concept

The JC3IEDM data model was designed to provide a **common set of data** for exchange within **Command and Control (C2) systems**, fully aligned with the concept of a **System of Systems (SoS)** [2]. The model describes all operational objects relevant to **disaster management operations**, including organizations, personnel, equipment, facilities, geographic features, and meteorological phenomena. It also provides a **graphical representation of the conceptual model** of JC3IEDM, showing the exchanged data, their relationships, and the associated business rules.

This information is critical for crisis management, as it captures the location, status, functions, and interactions of objects in a disaster area. The JC3IEDM data model can be considered a **meta-model**, offering a structured framework for the following elements:

- Objects and their properties;
- Past, present, and future situations, described through facts related to these objects;
- Past, present, and future activities involving these objects;
- Mechanisms for grouping data into information packages.

Key characteristics of the JC3IEDM model include:

- A sufficiently **generalized structure** to support joint operations across land, maritime, and air domains;
- Comprehensive description of all objects within an operation, including organizational hierarchy, personnel, equipment, assets, geographic conditions, and meteorological factors;
- Each object has **capabilities** for performing assigned functions, providing insight into its role and importance within the operation;
- Geographic positioning of objects is possible, with support for **geometric constructs** for planning, management, and operational tracking. Examples include state borders, corridors, protected zones, and other features relevant to commanders and their staff;
- Complex types are composed of specific **subtypes**, including tables for organizations, equipment, and personnel;
- Relationships between pairs of objects (relations) can be recorded, such as the association of organizational units, operational tasks, objectives, and operational entities;
- Metadata on the **information source, reporting time, and validity** of data is maintained.

The JC3IEDM data model supports **integration with distributed information technologies** [8, 9], including HLA (High Level Architecture), DDS (Data Distribution Service) [10], SOA (Service-Oriented Architecture), and SOAP (Simple Object Access Protocol). This flexibility establishes JC3IEDM as a key enabler for modern crisis management information systems.

4.2. Key Structural Elements of JC3IEDM

A fundamental element in the JC3IEDM specification is the **"entity"**, representing any distinguishable individual, location, object, event, or concept for which information will be recorded. These independent entities are those whose identification does not depend on any other entity [7]. All remaining entities are classified as **dependent entities**.

The independent entities and the relationships between them are illustrated in Figure 1. A dot at the end of a relationship indicates **"many"**. The relationships shown in the diagram are either **many-to-many** (solid line with two dots) or **one-to-many** (dashed line). Some relationships are **recursive**, reflecting entities that can relate to other instances of the same type. In the context of **crisis management operations**, these entities represent key operational components - personnel, equipment, facilities, and events—allowing the data model to support **situational awareness, operational planning, and coordination** across multiple agencies and organizational levels.

4. Selection of Technology for Access to the Database

In the standard approach to database access, a new connection to the database is opened and closed for each individual query. This requires the programmer to explicitly manage the connection lifecycle-opening the database connection, constructing the SQL query, executing it, and then closing the connection. In practice, issues often arise due to improperly closed connections, which can lead to inefficient use of system resources.

When applying such an approach to retrieve information from the **JC3IEDM database**, it is often necessary to construct complex and lengthy SQL queries. In large-scale systems, this process may become time-consuming and difficult to manage. Therefore, this approach is generally recommended only for relatively small systems, while in complex environments it may significantly complicate system maintenance and development.

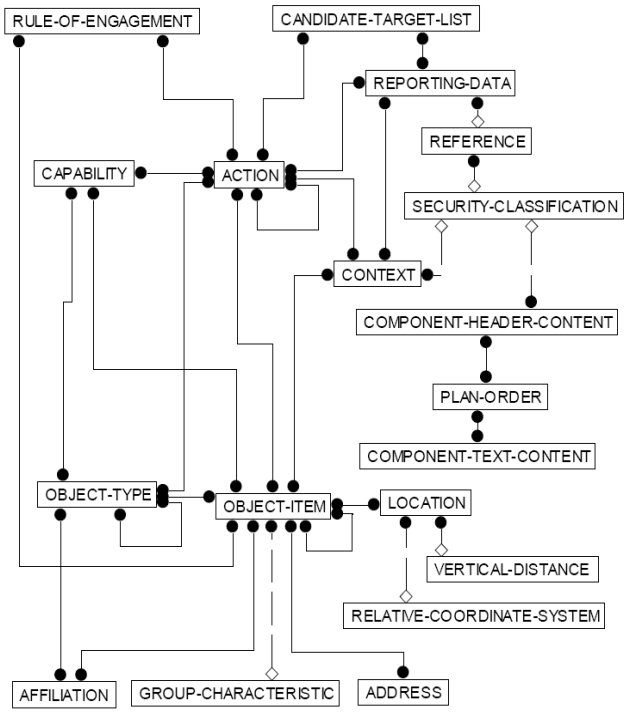


Fig. 1. Independent entities and the relationships between them

Object-Relational Mapping (ORM) represents an additional architectural layer within a software system. It is commonly used in systems developed with **object-oriented programming (OOP)**. This approach establishes a correspondence between the concepts of relational database modeling and object-oriented design through a **class-table mapping** [11]. The primary objective of ORM is to transform the complex hierarchical structure of application objects into a system of interrelated database tables containing different data types.

The complexity of this process depends on the specific objects and their implementation, but in general it can be divided into two main tasks: **storing object attributes** and **storing relationships between objects**. Unlike the direct database access approach, ORM enables interaction with the database in a way that returns **objects rather than raw database records**. The use of ORM provides several advantages:

- Faster development of the **database access layer**;
- Easier integration with the rest of the system code while preserving object-oriented paradigms such as **polymorphism**, since developers work with objects rather than raw data;
- Automatic generation of complex queries with minimal developer intervention;
- Support for **inheritance and hierarchical relationships** between objects;
- Efficient **caching mechanisms** that improve overall system performance.

In the context of **crisis management information systems**, these advantages are particularly important. Crisis response environments require the rapid processing and exchange of large volumes of operational data related to personnel, resources, infrastructure, and evolving situations. The use of ORM technologies facilitates efficient access to JC3IEDM -

based data structures and supports the development of scalable, interoperable applications capable of operating in **network-centric and multi-agency crisis management environments**.

The range of available **Object-Relational Mapping (ORM)** software solutions is quite extensive. However, many of these packages are either in an early stage of development and have not yet demonstrated sufficient stability and functionality, or they are relatively unpopular, which makes system development difficult due to limited documentation and community support. Within this study, the following technologies were analyzed: **LINQ, Entity Framework, and NHibernate**.

Language Integrated Query (LINQ) is a set of Microsoft .NET technologies that provides built-in query functionality similar to SQL. Unlike traditional SQL-based approaches, LINQ allows queries not only against relational databases but also against data from various sources, such as XML documents and in-memory collections. It is a powerful tool that resembles SQL-like languages used to process rows in database tables. LINQ can generally be divided into three main categories: **LINQ to Objects, LINQ to SQL, and LINQ to XML**. The technology focuses primarily on the method of constructing queries within the programming environment. However, according to some performance evaluations, the transformation of LINQ queries into SQL statements may introduce additional processing overhead [11].

Entity Framework is an object-oriented database access technology developed by Microsoft that allows developers to work at a higher level of abstraction. It represents an ORM solution integrated within the **.NET Framework**, enabling developers to manipulate database objects as application entities rather than directly managing database tables and SQL statements.

NHibernate is another suitable ORM solution for the .NET platform. As an open-source framework, it offers the advantage that potential issues can be addressed by the development community without depending solely on the response time of a large software vendor. NHibernate is essentially the .NET implementation of the widely adopted **Hibernate framework** for Java, inheriting many of the strengths of this well-established technology. In addition, numerous supporting tools are available that further facilitate the software development process.

Both **Entity Framework** and **NHibernate** represent reliable ORM solutions for software development on the .NET platform. In terms of functionality and performance, there are no significant differences between the two technologies. In this study, **NHibernate** was selected based on the following criteria:

- The software is **open source**, with comprehensive documentation and strong community support;
- Drivers are available for **all major database management systems**;
- A wide range of **additional development tools** exists to support and streamline the software development process.

In the context of **disaster management information systems**, the selection of an appropriate ORM technology is particularly important. Such systems must handle large volumes of operational data related to personnel, resources, infrastructure, and rapidly evolving situations. Efficient and flexible access to the **JC3IEDM database** supports interoperability between participating organizations and facilitates the development of scalable applications capable of operating in **distributed, network-centric crisis management environments**.

5. Object-Oriented Interface

This section presents the developed **intermediate Object-Oriented Interface (OOI)**, which enables direct access to the **JC3IEDM database** from various software applications. This approach facilitates the development of new applications as well as the adaptation of existing ones within a **network-centric information environment** that utilizes the JC3IEDM data model.

The Object-Oriented Interface is implemented using an **Object-Relational Mapping (ORM)** approach, specifically through the **NHibernate** framework. This design allows software components to interact with the JC3IEDM database through object-oriented structures rather than through direct SQL queries, thereby simplifying application development and improving system maintainability.

The core concept behind the OOI is based on the idea of **mapping** between application classes and the tables defined in the JC3IEDM data model. Once these mappings are defined, the ORM library manages the synchronization between objects in memory and the corresponding data stored in the JC3IEDM database. In NHibernate, these mappings are typically specified using **XML configuration files**.

From the perspective of **disaster management systems**, the implementation of such an interface provides several important advantages. Disaster response environments require the rapid integration of information from multiple heterogeneous systems operated by different organizations. The Object-Oriented Interface facilitates this integration by providing a **standardized and unified access layer** to the data structures. As a result, applications supporting **situational awareness, resource management, operational planning, and coordination among emergency response agencies** can access and exchange information more efficiently.

Furthermore, the use of an object-oriented interface enhances the **scalability and interoperability** of disaster management information systems. It enables different applications to operate within a **distributed, network-centric architecture**, ensuring that critical operational data-such as information about personnel, equipment, infrastructure, and evolving incidents-can be shared quickly and reliably among participating organizations.

Applications use **transactions** to process and manage objects. As a result of these transactions, the state of objects may change, and such changes must be reflected in the database. In the developed **Object-Oriented Interface (OOI)**, a list of objects whose state has been modified within a transaction is maintained. The recording of changes and the prevention of concurrent access are handled through an object of type **Session**.

In this section, an example of the **ObjectItemCapability** object from the developed Object-Oriented Interface for accessing the **JC3IEDM database** through **NHibernate** is presented. Initially, a .NET class named **ObjectItemCapability** is created (Figure 2), which corresponds to a specific table in the JC3IEDM database. The data in memory are represented by the fields of this class, enabling the retrieval, modification, and updating of the corresponding records.

In the context of **disaster management information systems**, the use of transactions is particularly important because operational data-such as information about available resources, capabilities of response units, equipment status, and operational tasks-are continuously updated during emergency response activities. The transactional approach ensures **data consistency, reliability, and integrity**, which are critical for maintaining accurate **situational awareness and coordination among participating organizations**. Moreover, the session-based mechanism helps prevent conflicts when multiple applications or agencies simultaneously access and update shared operational data in a **distributed, network-centric crisis management environment**.

```
public class ObjectItemCapability : Generic.Entity
{
   // private fields

    private AbstractObjectItem objectItem { get; set; }
    private AbstractCapability capability { get; set; }
    private MissionPrimacyCode missionPrimacyCode;
    private QuantityNonNegative12_3 quantity { get; set; }
    private AbstractReportingData reportingData { get; set; }

    // default constructor

    public ObjectItemCapability()...

    // getter & setter methods

    public AbstractObjectItem getObjectItem()
    {
        if (objectItem == null)
        {
            throw new NullValueException(
                "ObjectItemCapability.objectItem");
        }
        return objectItem;
    }
```

Fig. 2. NET ObjectItemCapability class

The next step is to implement the **mapping** between the SQL table and the **ObjectItemCapability** class (Figure 3). This is achieved through a **mapping configuration file**. In the developed interface, each class is associated with a corresponding XML mapping file following the naming convention **ClassName.hbm.xml**.

In the context of **disaster management information systems**, this mapping mechanism ensures a consistent correspondence between operational data stored in the **JC3IEDM database** and the application-level objects used by crisis management software. Such an approach facilitates reliable access to information related to **operational capabilities, available resources, personnel, and infrastructure**, which are essential for effective coordination during emergency response operations. By maintaining a clear and structured mapping between database entities and application objects, the system supports efficient data exchange and integration across multiple applications within a **distributed, network-centric crisis management environment**.

```
<?xml version="1.0" encoding="UTF-8"?>
- <hibernate-mapping xmlns="urn:nhibernate-mapping-2.2">
  - <class lazy="false" abstract="false" table="OBJ_ITEM_CAPAB"
    name="Model.Entity.ObjectItemCapability, Model.Entity">
    - <composite-id name="ID"
      class="Model.Entity.ObjectItemCapabilityIdentifier, Model.Entity">
        <key-property name="ObjectItemId" type="decimal(20,0)"
          column="OBJ_ITEM_ID"/>
        <key-property name="CapabilityId" type="decimal(20,0)"
          column="CAPAB_ID"/>
        <key-property name="INDEX" type="decimal(20,0)"
          column="OBJ_ITEM_CAPAB_IX"/>
      </composite-id>
      <version name="Version" type="timestamp" column="Version"
        access="field.pascalcase-underscore"/>
      <many-to-one name="objectItem"
        class="Model.Entity.AbstractObjectItem, Model.Entity"
        column="OBJ_ITEM_ID" cascade="none" update="false"
        insert="false"/>
      <many-to-one name="capability"
        class="Model.Entity.AbstractCapability, Model.Entity"
        column="CAPAB_ID" cascade="none" update="false" insert="false"/>
      <property name="INDEX" type="decimal(20,0)"
        column="OBJ_ITEM_CAPAB_IX" update="false" insert="false"/>
      <property name="MissionPrimCode" type="string"
        column="MSN_PRIMACY_CODE" not-null="false" length="6"/>
    - <component name="quantity"
      class="Model.Simple.QuantityNonNegative12_3, Model.Simple">
        <property name="Value" type="decimal(12,3)" column="QTY"/>
      </component>
      <many-to-one name="reportingData"
        class="Model.Entity.AbstractReportingData, Model.Entity"
        column="RPTD_ID" cascade="none" not-null="false" fetch="select"
        not-found="ignore"/>
    </class>
</hibernate-mapping>
```

Fig. 3. Mapping - SQL table and the ObjectItemCapability class

6. Conclusion

The integration of **JC3IEDM data exchange models** into disaster management systems supports the realization of the objectives for interoperability among **Command and Control (C2)** information systems. JC3IEDM establishes a common approach for describing the information exchanged between C2 systems. This provides:

- A **standard for interoperability** among C2 information systems;
- A **common interface** between heterogeneous systems;
- A **shared understanding of exchanged information** across different international organizations and levels of command.

The use of a common data model such as JC3IEDM facilitates **timely information exchange, improved situational awareness, and more effective coordination of resources and response actions**.

The proposed **Intermediate Object-Oriented Interface** provides direct access to the JC3IEDM database from various software applications. This enables the development of new applications as well as the adaptation of existing ones within a **network-centric information environment**. In the context of disaster management, such an approach supports the creation of **integrated decision-support systems**, improves the reliability and accessibility of operational data, and enhances the overall effectiveness of **multi-agency and multinational crisis response operations**.

References

1. Chigani, A., Arthur, J., The implications of network-centric software systems on software architecture: A critical evaluation, Proceedings of the 45th Annual Southeast Regional Conference, Winston-Salem, North Carolina, USA, March 23-24, 2007, 70-75, ISBN 978-1-59593-629-5.
2. Systems Engineering Guide for Systems of Systems, Version 1.0, August 2008, Washington, DC: ODUSD(A&T)SSE, 2008.
3. Mitrev, D., Saltirov, D., Uzunov, R. Management of Complex Systems: A Distance Learning Textbook. University of Shumen National Security University of Shumen (in Bulgarian)
4. UK DMWG: JC3IEDM Overview - The Joint C3 Information Exchange Data Model Overview. Edition 3.1a. (2007., https://brainsciencesjournal.org/2007.JC3IEDM.pdf
5. Serna, M., Beltrán, M., Guzmán, A., "Achieving Interoperability in Military Simulation Applications: Defining Models for HLA Utilization",International Journal of Modeling and Optimization, 2012.
6. UK DMWG: JC3IEDM Overview - The Joint C3 Information Exchange Data Model Overview. Edition 3.1a. (2007., https://brainsciencesjournal.org/2007.JC3IEDM.pdf,
7. Gerz, M., Meyer, O., "Defining C2 semantics by a platform-independent JC3IEDM", International Journal of Intelligent Defence Support Systems, 2011.
8. IEEE, P 1516, "Standard for Modeling and Simulation (M&S) High Level Architecture (HLA) - Framework and Rules", February 2000.
9. Kuhl F., R. Weatherly, and J. Dahmann, "Creating Computer Simulation Systems: An Introduction to the High Level Architecture", Prentice Hall PTR, 1999, ISBN 0130225118.
10. Kirov, G., Stoyanov, V., Lazarov, B.: Abstract model of an object-oriented layer for distributed systems based on the DDS standard. In: Proceedings of the 12th International Conference on Computer Systems and Technologies, pp. 75-81 (2011)

The Institutionalization of Korean Studies in Bulgaria: From Foundational Drivers to Future Prospects

Hwan Joung James Kang

University of Plovdiv "Paisii Hilendarski", University of Veliko Tarnovo "St. Cyril and St. Methodius," Bulgaria
jameshjkang@uni-plovdiv.bg

Abstract. This paper examines, from a diachronic perspective, the development of Korean Studies within the Bulgarian higher education system, from its institutional inception in 1995 to its current status and future prospects. The paper identifies seven key driving forces behind the establishment and expansion of Korean Studies in Bulgaria - state support, faculty and students, diplomatic engagement and university leadership, Korea's soft power and smart power, the European context and the Bulgarian perspective, the academic and ethical significance of Hangeul, and the normative philosophy of Hongik Ingan. The geographical and historical context of post-Cold War Bulgaria was selected to highlight the social impact and institutional achievements of Korean Studies. In addition, the paper argues that Korean Studies in Bulgaria was an academic fostering cultural exchange and civic participation. On this basis, the article suggests that the future consolidation of Korean Studies in Bulgaria will depend on continued institutional support, strengthened academic networks, interdisciplinary cooperation, and the deeper transmission of Hongik Ingan as an educational and ethical principle.

Keywords: Korean Studies; Bulgaria; higher education; soft power; Hangeul; Hongik Ingan

1. Introduction

Over the past three decades, the globalization of Korean Studies has accelerated significantly, and its expansion has extended beyond East Asia to various regions of the world. In its initial phase, Western Europe played a central role in the institutionalization of Korean Studies, but Eastern Europe also gradually emerged as a new space of institutional innovation and cultural reception. In this context, Bulgaria represents a particularly noteworthy case. Despite limited resources and a relatively short institutional history, Bulgaria succeeded in establishing an institutional foundation for Korean Studies within its own higher education systemand gradually expanding it into a national academic and cultural network.

The objectives of this study are twofold. First, it seeks to identify the principal driving forces that made the development of Korean Studies in Bulgaria possible. Second, it aims to present the broader academic and social implications that this development has for both Bulgaria and Korean Studies in Europe as a whole.

2. The History and Institutional Growth of Korean Studies in Bulgarian Higher Education Institutions

The beginnings of Korean Studies in Bulgaria can be traced to the establishment of diplomatic relations between the Republic of Korea (Korea) and the Republic of Bulgaria on March 23, 1990. Later, in 1995 the Bulgarian President Zhelyu Zhelev had an official

delegation to Korea (April 2-5, 1995) and in parallel with the proposal to establish Bulgarian Studies at Hankuk University of Foreign Studies in Seoul, it was proposed that Korean Studies be established at Sofia University. It was precisely within this political and diplomatic context that Korean Studies at Sofia University was officially inaugurated in 1995.

President Zhelyu Zhelev (1935-2015) was the first Bulgarian president elected by direct popular vote after the systemic transformation. He maintained a long-standing connection with Choi Dong-Seong, a North Korean student in Bulgaria, who later assisted his as an interpreter during the 1995 official delegation to South Korea. In 1992 Korean language had already been preliminarily introduced and conducted within the Center for Eastern Languages and Cultures at Sofia University.

Later that same year, Sofia University opened the Korean Studies program with approximately ten students. In 2003, a separate building for the Center for Korean Studies was established apart from the building of the Center for Eastern Languages and Cultures. In this process, Korea's government financing amounting to 300 000 USD enabled the Korean Studies to secure a permanent institutional base equipped with its own library and the necessary academic resources. In addition, the program deepened its institutionalization through the sustained support of the Korea Foundation (KF), the Academy of Korean Studies (AKS), and the Sejong Institute Foundation (SIF). The establishment of the King Sejong Institute at Sofia University in 2013 constituted yet another symbolic event demonstrating that both the institutionalization and the social expansion of Korean Studies had entered a new stage.

In 2010, Sofia University elevated Korean Studies to the status of an independent department. At that time, the department had secured seven full-time faculty members and the number of newly admitted students was approximately twenty per year. This indicated that Korean Studies had achieved a recognized level of academic maturity, extending beyond the level of a simple language training unit. By way of comparison, Chinese Studies at Sofia University began as a language course in 1953 and was structured as an official academic program in 1991, yet it was not elevated to the status of an independent department until 2018. Similarly, Japanese Studies began in 1968 in the form of language courses and was recognized as a regular university specialty in 1990, but, like Chinese Studies, it was not reorganized as an independent department until 2018. From this comparative perspective, Korean Studies may be evaluated as having grown into an independent department within a relatively short period of time and, furthermore, as having developed into a central hub of Korean Studies equipped with a comprehensive higher education structure encompassing bachelor's, master's, and doctoral levels.

Thereafter, Korean Studies began to spread to other universities as well. The cooperation between the Korean Embassy in Bulgaria and the Department of Korean Studies at Sofia University gradually led to the institutional establishment of specialty of Korean Studies at Plovdiv University in 2014. At Veliko Tarnovo University, Korean language courses was introduced in elective form in the 2012/2013 academic year and, from 2013/2014 onward, Korean Studies was incorporated into the Applied Linguistics curriculum.

Currently, courses in Korean language are offered at New Bulgarian University, and Korean-related electives are also conducted at Trakia University. In addition, from the autumn of 2026, Sofia University plans to open a new Korean Studies bachelor's program at its Burgas branch. Korean Studies has gradually gained academic expansion among Bulgarian universities.

3. Expansion of Korean Studies in Bulgaria

The influence of Korean Studies in Bulgaria started expanding not only in tertiary education but also in primary and secondary schools, thereby broadening the age range of Korean language learners. In 2011, the 18th Secondary School "William Gladstone" in Sofia, Bulgaria, became the first in Europe to integrate Korean into the regular national state curriculum as a primary foreign language. The program subsequently got extended to the primary school levels and Korean was studied as a first foreign language during the whole period of primary and secondary education. At the same time, Korean language club activities developed in secondary schools in Sofia, Vratsa, Varna, Burgas, and Veliko Tarnovo. With the establishment of the King Sejong Institute in 2013 in Sofia, public access to Korean language and Korean culture education increased even further. In addition, the King Sejong Institute annually administers the TOPIK examination, that is, the Test of Proficiency in Korean.

Furthermore, private Korean language teaching businesses and club activities have spread across various regions, including Sofia, Varna, Burgas, Plovdiv, Stara Zagora, Vratsa, and Gabrovo, whereby Korean language and Korean culture have gradually become part of the everyday experience of Bulgaria's younger millennials and generation Z.

At the same time, academic cooperation has also steadily expanded. Sofia University has established cooperative relations with Hankuk University of Foreign Studies and Sungkyunkwan University, while Veliko Tarnovo University has deepened academic exchange through its partnership with Kangnam University. These universities have made student and faculty mobility possible through the Erasmus+ program and bilateral exchange agreements, thereby providing a foundation for more advanced joint research and intercultural exchange. Plovdiv University has also contributed to the spread of Korean Studies in Southern Bulgaria through the programs including Korean language and through the organization of regional events on Korean culture. All these progressions clearly demonstrate that Korean Studies in Bulgaria has developed from a small-scale program in a singular university into programs in multiple universities across the country and a national network in academia.

4. The Seven Driving Forces of Development

The success of Korean Studies in Bulgaria may be explained through the interaction of seven key driving forces, each of which reinforces the others and thereby sustains the overall structure of development.

4.1. State Support

In the development of Korean Studies in Bulgaria, support from both the Korean and the Bulgarian states has been of decisive significance. On the Korean side, government entities such as KOICA (Korea International Cooperation Agency), the Korea Foundation (KF), the Academy of Korean Studies (AKS), the Global Korea Scholarship (GKS), and the Sejong Institute Foundation (SIF) have played central roles in providing financial support, educational materials, exchange opportunities, and the long-term cultivation of human resources. The Korean government has continuously provided support int he form of financing and educational materials to the Department of Korean Studies at Sofia University through Korean Studies-related programs such as the Seed Program and the Core University Program, and through seminars, forums, literary translation initiatives, and other academic and cultural programs. In addition, professors and students have regularly been invited to Korea, which has constituted an investment in the long-term training of specialists in Korean Studies. Annually

around ten students participate in the Global Korea Scholarship (GKS) program and enroll in Korean universities as scholarship recipients.

In addition, the Embassy of the Republic of Korea in Bulgaria has continually acknowledged the development and Korean Studies at Sofia University by organizing meetings between the Department and high-ranking Korean representatives during their visits in Bulgaria.

This tendency became particularly pronounced during the tenure of Prof. Aleksander Fedotov (1956-2018). While serving as the head of the Center for Korean Studies from 2001 to 2007, he also held the position of Vice Rector for International Affairs at Sofia University.

Furthermore, as mentioned above, after the opening of the Korean language program at the 18th School in Sofia, the Korean government continued the provision of technical equipment and educational materials. As with Sofia University, the Korean Embassy organized meetings between school's leadership team and high-ranking figures from Korea during their official visits in Bulgaria. In addition, the Korean stakeholders were involved in dispatching native Korean language teachers into the school. Public access to information about Korea was facilitated by the Korean government with the establishment of the "Korean Corner" at Sofia City Library, a space dedicated to Korean books and literary events, and the continuous replenishment of a comprehensive collection of literature and multimedia.

Bulgarian stakeholders have equally supported the development of Korean studies. The Bulgarian state officially recognized Korean Studies as a legitimate academic and cultural field, thereby facilitating its stable integration into the national higher education system. In addition, the inclusion of Korean language within state matriculation examinations for secondary education – as a rare elective language - demonstrates its transition from a peripheral interest to an institutionally integrated academic discipline. It was precisely this multilayered support that reduced the initial institutional risks, stabilized growth, and conferred upon Korean Studies both academic legitimacy.

4.2. Faculty and Students

The advancement of Korean Studies at Sofia University is a result of the collective commitment of both faculty and student body. Prof. Choi Gwon-jin and Prof. Alexander Fedotof provided the academic groundwork through research and curriculum development, and with the joint effort of Prof. Kim So-young they enabled the systematic education of Bulgaria's first generation of Korean Studies specialists.

Over time, several alumni transitioned into faculty and research roles, ensuring academic continuity and fostering a scholarly community deeply integrated in the local context. This evolution illustrates the transformation of Korean Studies in Bulgaria from a discipline dependent primarily on external support into a self-sustaining academic field. In addition, a Center for Korean Language and Culture was established at Veliko Tarnovo University, where graduates now serve as lecturers alongside a Korean scholar Prof. James Kang. Furter, at Plovdiv University, initiatives such as the annual "Day of Korea" and seasonal student camps - spearheaded by faculty such as assistant professor Cathy Kang - underscore the field's maturity. Collectively, these developments demonstrate that Korean Studies in Bulgaria has established a robust internal mechanism for academic succession.

4.3. Diplomatic Engagement and University Leadership

Diplomatic initiatives and strategic institutional leadership represent pivotal drivers in the advancement of Korean Studies within the Bulgarian academic landscape. The Embassy of the Republic of Korea in Sofia has consistently facilitated high-level academic engagements

and cultural symposia, thereby augmenting the discursive visibility of Korean Studies across the broader Bulgarian socio-cultural sphere. Furthermore, by fostering robust networks between Bulgarian and Korean scholars, the Embassy has successfully integrated the discipline into its core diplomatic priorities.

Concurrently, the rectoral administration of Sofia University has functioned as a critical institutional patron, allocating requisite resources and ensuring the program's public visibility. The attainment of departmental autonomy and the subsequent institutionalization of the field would have been improbable without the convergence of diplomatic support and university governance. The multidimensional collaboration has provided Korean Studies with the structural stability and academic credibility necessary to emerge as a resilient research field, insulated shifting geopolitical climates.

4.4. Korea's Soft Power and Smart Power

The burgeoning interest in Korean Studies within the Bulgarian academic landscape is fundamentally rooted in Korea's multifaceted emergence as a global power. Historically, this trajectory began with Korea's rise as an economic and technological powerhouse, defined by the industrial "Miracle on the Han River" and the global dominance of its electronics and automobile manufacturing sectors. In the Bulgarian context, this foundational economic strength provided the platform for the subsequent global diffusion of soft power through the "Korean Wave" (or *Hallyu*). Initially driven by acclaimed cinema, K-pop, and the lifestyle appeal of K-food and K-beauty, this cultural capital has profoundly reshaped the socio-cultural imaginary of Bulgaria's younger demographic, serving as a primary catalyst for the systematic pursuit of Korean Studies.

In the contemporary era, the scope of Korean influence has undergone a strategic pivot, integrating hard power and institutional soft power into its global profile. This is manifested in the international competitiveness of K-defense and K-shipbuilding, alongside the export of governance models through K-education and K-democracy.

Recently, a new dimension has been added to this in the form of Korea's smart power. Facilitated by the Ministry of Science and ICT, the state has formally set the strategic objective to achieve "AI G3" status, aiming to become one of the preeminent global leading powers in artificial intelligence. To this end, the country is developing national AI infrastructure, local generative AI models, AI semiconductors, resources for large-scale computation, and the widespread application of AI across diverse industrial sectors and the public sector. A pivotal milestone was the recent signing of Letter of Intent with six United Nations agencies to collaborate on establishing a Global AI Hub in Korea.

Consequently, the field of Korean Studies in Bulgaria is undergoing an essential academic broadening. While the discipline's initial attraction was rooted in Korea's resilience and contemporary soft power, it must now expand its horizons to encompass the nation's socio-technical transformations, emerging role in international governance, and future industrial trajectories. It is precisely this convergence of soft and smart power that not only attracts students to the field but also provides the structural depth necessary to sustain their long-term academic engagement. By integrating these multifaceted dimensions, Korean Studies secures its position as a rigorous, multidimensional subject of inquiry, resilient to shifting popular trends.

4.5. The European Context and the Bulgarian Perspective

The evolution of Korean Studies in Bulgaria is inextricably linked to the broader European geopolitical reconfiguration following the end of the Cold War. In the wake of

Mikhail Gorbachev's (1931-2022) declaration of *Perestroika* in 1986 and the seminal events of November 1989, Europe turned toward the search for new partners for cooperation, Eastern European countries also began to rethink their relations with Asia. Bulgaria participated in this process while simultaneously embracing new languages and cultures in the spirit of the cultural maxim that "a new language creates new wealth."

Consequently, the study of the Korean language has been reconceptualized as a high-value strategic asset, transcending its role as. A medium for cultural enrichment. In contemporary landscape, proficiency in Korea is viewed as a significant investment in human capital, offering expanded prospects within the global academic labor market and diverse professional sectors. Furthermore, the study of Korean functions as a critical instrument of track-two diplomacy, facilitating deeper institutional and economic integration between Bulgaria, the European Union, and Korea. This strategic vision has been further institutionalized through the European Union's Erasmus+ program and a robust network of bilateral university agreements. Notable examples include Sofia University which has cultivated sustainable cooperative relations with Hankuk University of Foreign Studies and Sungkyunkwan University, and Veliko Tarnovo Universit, which has similarly deepened its academic ties with Kangnam University. By intensifying bilateral mobility of students and faculty, these partnerships have established a permanent foundation for advanced intercultural discourse and joint research, ensuring the long-term vitality of Korean Studies.

4.6. The Beauty, Scientific Character, and Spirit of the Korean Alphabet "Hangeul"

Hangeul (originally *Hunminjeongeum*), created and promulgated in 1446 by King Sejong the Great. It is a unique writing system on a global scale because its origin, principles of creation, and structural scientific character are clearly documented. This transparency has garnered significant International academic interest. In 1989, UNESCO established the King Sejong Literacy Prize to honor this spirit of literacy, and in 1997, the *Hunminjeongeum Haerye* was inscribed in the UNESCO Memory of the World Register. This recognition underscores that the creativity and importance of Hangeul as a vital component of the cultural heritage of humankind and global literacy. *Hunminjeongeum Haerye* is the 1446 explanatory edition of the Korean script promulgated by King Sejong, containing both the royal preface and scholarly commentaries that explain the purpose, sounds, and principles of the newly created alphabet. It is widely regarded as an extraordinary documentary record because it clearly preserves the original rationale and design of Hangeul.

Hangeul is a unique writing system composed of a highly logical alphabet with individual letters representing specific sounds and with organized syllabic blocks. The success of Korean language education globally, exemplified by its development in Bulgaria – is largely due to the scientific structure and phonological transparency. Its scientific structure, phonological transparency, and relative ease of acquisition lower the threshold of entry for learners and enable them to move comparatively quickly from purely linguistic training to content-oriented disciplines related to history, literature, and society. Moreover, in the digital age, the structural clarity and systematicity of Hangeul make it particularly attractive for fields such as computational linguistics and digital humanities.

The significance of Hangeul, however, is not exhausted by its linguistic efficiency alone. At its foundation lies the *aemin* spirit, that is the a "spirit of love for the people" - King Sejong's compassion and care for ordinary people. It is designed specifically to empower those who lacked the means to learn complex classical scripts. Viewed through this lens, Hangeul is not merely as a technical writing system but as a symbol of human-centered education. This ethical dimension, in turn, finds an especially strong resonance in the Bulgarian educational

and cultural context, in which language education is often conceived as a form of social contribution.

4.7. The Philosophy of Hongik Ingan

One profound, yet often overlooked, driver in the advancement of Korean Studies in Bulgaria is the philosophy of *Hongik Ingan*. Identified as a foundational philosophy of the Republic of Korea and explicitly codified in Article 2 of the Education Act, the term translates to "broadly benefit the human world." This principle imbues Korean Studies with universal significance, shifting the paradigm of education from a mere means of individual success to a framework for service and contribution to the collective welfare.

The ethical shift became increasingly visible as Korea emerged as a global economic power and continued to develop its international image, it sought to interpret itself as a country that contributes to the global community through culture, education, technology, and a commitment to two-way exchange. Consequently, philosophy of *Hongik Ingan* imparts to Korean Studies not merely a cognitive, but also a moral orientation. This ethos aligns with the seminal vision of As Kim Gu who famously posited in his essay *My Wish* that true national greatness is predicated not on military or economic hegemony, but on the cultivation of a "noble culture". Kim Gu (1876-1949) was a leading Korean independence activist and the president of the Provisional Government of the Republic of Korea. He emphasized that true national strength lies not in military or economic power but in cultural and moral values. He envisioned Korea as a nation that contributes to the world through its rich culture, humanity, and ethical ideals. He argued that culture a force capable of fostering a personal and altruistic sense of "happiness" and aspired that Korea would behold a culture beautiful and beneficial to others. This is directly connected with the Korean cultural ideal that encounters the world through Korean Studies.

Within the Bulgarian academic landscape, this philosophy has been institutionalized as both a pedagogical cornerstone for faculty and a moral heuristic for students. By internalizing the tenets of *Hongik Ingan,* students are encouraged to synthesize academic proficiency with ethical empathy and social responsibility. In practice, this is manifested through a strategic integration of scholarship and civic engagement, exemplified by collaborative public translation projects, the preservation of cultural heritage, and partnerships with local communities. Consequently, the adoption of *Hongik Ingan* ensures that Korean Studies in Bulgaria transcends the sole transmission of empirical knowledge, and it facilitates an entry into the discourse of shared human values, bridging regional academic pursuits with global imperatives for peace, international cooperation, and sustainable development.

5. Future Prospects

Several primary trajectories emerge regarding the future development of Korean Studies in Bulgaria. First, the foundational drivers that have historically sustained the field's evolution are projected to maintain their momentum, thereby ensuring institutional stability and continued growth. The existing institutional infrastructure, coupled with accumulated pedagogical and administrative experience, constitutes significant intellectual capital. The expansion of Korean Studies across Sofia University, Plovidiv University, Veliko Tarnovo University, and the newly established Korean language program at a regional branch of Sofia University in Burgas indicates that the institutional and geographical footprint of the Korean Studies will continue to broaden.

Second, the deepening of the academic foundation of Korean Studies is contingent upon the fortification of transnational networks and the formation of joint research clusters between

Bulgarian and Korean institutions. By prioritizing interdisciplinary fields – such as digital humanities, translation studies, and comparative social studies – the discipline can transcend the traditional boundaries of area studies. This transition is essential for integrating Korean Studies into broader international academic discourses. Furthermore, Korea's strategic advancements in artificial intelligence and digital transformation offer a solid conceptual background for pioneering future-oriented research agendas.

Third, the long-temr sustainability of Korean Studies will increasingly depend on the degree to which the philosophy of *Hongik Ingan* is effectively integrated into the academic development of faculty and students. It is imperative that this philosophy is embraced not only as an abstract universal ethic, but as a practical heuristic guiding individual conduct and social engagement. By internalizing these values, Korean Studies in Bulgaria will be able to preserve its academic dynamism and serve as a catalyst for the development of the Bulgarian civil society.

6. Conclusion

The evolution of Korean Studies in Bulgaria serves as a pragmatic case study of how a modest institutional initiative – commencing in 1995 with a cohort of around ten student - can transform into a robust academic and cultural network of both national and European significance. Its developmental trajectory has been shaped by the convergence of seven primary drivers: institutional state support, the proactive engagement of faculty and students, diplomatic engagement and university leadership, the strategic application of Korea's soft power and smart power, the integration of European and Bulgarian perspectives, the aesthetic and linguistic appeal of Hangeul, and the overarching philosophy of *Hongik Ingan*.

These constitutive forces elucidate not only the historical expansion of the discipline of Korean Studies but also its emergence as a critical conduit and a platform for substantive civic contribution. By transcending the conventional boundaries of linguistic and cultural pedagogy, Korean Studies in Bulgaria has evolved into a pivotal socio-academic sphere. In this domain, bilateral relations are fortified, human values are institutionalized, and intercultural discourse is systematically advanced.

If the current momentum of sustained support, the expansion of transnational network, and the revitalized commitment to the philosophy of *Hongin Ingan* are maintained, Korean Studies in Bulgaria is uniquely positioned for long-term sustainability and possesses the necessary intellectual and institutional foundations to flourish as a vital field of scholarship for future generations.

References

1. Borisova, Tsvetelina. "A Case Study on the Evolution of K-pop and K-drama Related Fandom in Bulgaria." *Manas: Studies into Asia and Africa* 6, no. 1 (2020): 22–31.
2. Choi, Gwon Jin. "Teaching the Korean Grammar in East Europe." *The Education of Korean Language* 16 (2005): 351–388.
3. Choi, Moon-hyoung. "The Concept of God in the Idea of Hongik Ingan." *The Review of Korean Studies* 12, no. 3 (2009): 177–205.
4. Dimitrova, Aneta, and Yana Mancheva, eds. In *The Focus of Korean Language, Culture and Society: Proceedings of the International Conference on Korean Studies, Sofia, May 26–27, 2023*. Sofia: St. Kliment Ohridski University Press, 2023.
5. Fedotoff, Alexander, and Kim So Young, eds. *Korea and Eastern Europe: Exploring the Past.* Sofia: St. Kliment Ohridski University Press, 2017.

6. Fedotoff, Alexander, and Kim So Young, eds. *On Religious and Cultural Contents in Korea and Bulgaria: Proceedings of the International Conference on Korean Studies, Sofia, November 16–17, 2017*. Sofia: Kamea Group Ltd., 2017.
7. Fedotov, Alexander. "The Status of Korean Studies in Bulgaria." In *White Paper on Korean Studies Abroad*, 484–494. 2007.
8. Kang, James. *Korea and Bulgaria: From Diplomatic Recognition to Prospective Cooperation, 1990–2014*, ISBN: 978-619-91739-2-3
9. Karteva-Dancheva, Svetla, and Kim So Young, eds. *Globalization in Korea: On the Occasion of the 30th Anniversary of the Establishment of Diplomatic Relations between the Republic of Bulgaria and the Republic of Korea*. Sofia: St. Kliment Ohridski University Press, 2020.
10. Kim, Hun Tae. "Current Status of Korean History Classes at European Universities and Examples of Educational Material Use." *CEFIA Newsletter* 199 (February 2021).
11. Kim, Soyoung. "Current Status and Prospects of Korean Studies in Bulgarian Higher Education." *Journal of Korean Culture* 65 (2024): 119–146.
12. Lee, Geun. "A Soft Power Approach to the 'Korean Wave.'" *The Review of Korean Studies* 12, no. 2 (2009): 123–137.
13. Muzaffar, S. A. "Public Diplomacy: Reappraising the South Korean Case." *Korea Journal* 57, no. 3 (2017): 83–113.
14. Sotirova, Irina, Svetoslava Peycheva, Alexander Fedotov, Miroslava Zaburtova, Svetla Karteva-Dancheva, Yana Mancheva, and Raina Beneva. *Religions in Korea: Tradition, Modernization and Globalization*. Sofia: St. Kliment Ohridski University Press, 2022.

Internet References

1. 18th Secondary School "William Gladstone." "Korean Language." https://18sou.net/%D0%BA%D0%BE%D1%80%D0%B5%D0%B9%D1%81%D0%BA%D0%B8-%D0%B5%D0%B7%D0%B8%D0%BA/. Accessed April 11, 2026.
2. APEC. *APEC Artificial Intelligence (AI) Initiative (2026–2030).* https://www.apec.org/meeting-papers/leaders-declarations/2025/2025-apec-leaders--gyeongju-declaration/apec-artificial-intelligence-%28ai%29-initiative-%282026-2030%29. Accessed April 11, 2026.
3. 32nd APEC Economic Leaders' Meeting 2025. *APEC Artificial Intelligence (AI) Initiative (2026–2030).* PDF file.
4. Academy of Korean Studies (AKS), CEFIA. "Korean Language Education and Korean Studies in Bulgaria." https://www.aks.ac.kr/ikorea/index.do. Accessed April 11, 2026.
5. King Sejong Institute. https://www.ksif.or.kr/intro.do. Accessed April 11, 2026.
6. Korea Foundation. "Scholarship / Support Materials for Korean Studies Abroad." https://www.kf.or.kr/kfEng/main.do. Accessed April 11, 2026.
7. Korea Foundation. "Scholarship / Support Programs for Korean Studies Abroad." https://www.uni-sofia.bg/index.php/eng/the_university/faculties/faculty_of_classical_and_modern_philology/departments/korean_studies. Accessed April 11, 2026.
8. Korea.net. "Launch of the First-ever Regular Korean Language Class in Europe." https://www.korea.net/Government/Briefing-Room/Press-Releases/view?articleId=668&type=O. Accessed April 11, 2026.
9. Ministry of Science and ICT, Republic of Korea. "Blueprint for Korea's Leap to Become One of the Top Three AI Powerhouses (AI G3)." https://www.msit.go.kr/eng/bbs/view.do?bbsSeqNo=42&mId=4&nttSeqNo=1037&sCode=eng. Accessed April 11, 2026.
10. Ministry of Science and ICT (MSIT), Republic of Korea. "National AI Strategy Policy Directions / Blueprint for Korea's AI Innovation to Achieve AI G3 Status." https://www.msit.go.kr/eng/bbs/view.do?bbsSeqNo=42&mId=4&mPid=2&nttSeqNo=1040&sCode=eng. Accessed April 11, 2026.
11. New Bulgarian University. "Korean Language – Level A1." https://uppo.nbu.bg/bg/ezikovo-obuchenie/chuzhdi-ezici/korejski-ezik-nivo-a1. Accessed April 11, 2026.
12. Plovdiv University "Paisii Hilendarski." "Bulgarian Language and Korean Language." https://uni-plovdiv.bg/pages/index/495/. Accessed April 11, 2026.

13. Plovdiv University "Paisii Hilendarski." "Day of Korea at Plovdiv University 'Paisii Hilendarski.'" December 5, 2023. https://slovo.uni-plovdiv.bg/%D0%B4%D0%B5%D0%BD-%D0%BD%D0%B0-%D0%BA%D0%BE%D1%80%D0%B5%D1%8F-%D0%B2-%D0%BF%D1%83-%D0%BF%D0%B0%D0%B8%D1%81%D0%B8%D0%B9-%D1%85%D0%B8%D0%BB%D0%B5%D0%BD%D0%B4%D0%B0%D1%80%D1%81%D0%BA%D0%B8/. Accessed April 11, 2026.
14. Sofia University "St. Kliment Ohridski." "Department of Korean Studies." https://www.uni-sofia.bg/index.php/eng/the_university/faculties/faculty_of_classical_and_modern_philology/departments/korean_studies. Accessed April 11, 2026.
15. Sofia University "St. Kliment Ohridski." "Faculty Members of the Department of Korean Studies." https://www.uni-sofia.bg/index.php/eng/the_university/faculties/faculty_of_classical_and_modern_philology/departments/korean_studies/lecturers. Accessed April 11, 2026.
16. Trakia University – Stara Zagora. "Department of Social Sciences and Business Language Training." https://uppo.nbu.bg/bg/ezikovo-obuchenie/chuzhdi-ezici/korejski-ezik-nivo-a1. Accessed April 11, 2026.
17. UNESCO. "UNESCO King Sejong Literacy Prize Winners on Study Tour to the Republic of Korea." https://www.unesco.org/en/articles/unesco-king-sejong-literacy-prize-winners-study-tour-republic-korea. Accessed April 11, 2026.
18. Veliko Tarnovo University "St. Cyril and St. Methodius." "Center for Korean Language and Culture." https://www.uni-vt.bg/bul/pages/?page=4181&zid=3. Accessed April 11, 2026.
19. Veliko Tarnovo University "St. Cyril and St. Methodius." "Korean Language as a Second Foreign Language." https://www.uni-vt.bg/bul/pages/?page=2669&zid=3. Accessed April 11, 2026.

Principles of Developing Secure Software Applications

Rosen Kirilov

University of National and World Economy, Sofia, Bulgaria
rkirilov@unwe.bg

Abstract. The report examines the issues of developing secure software applications. The specifics of each of the development stages are presented in accordance with the requirements of ISO 27001. Based on the analyses performed, some principles for developing software platforms and applications are formulated. Finally, conclusions are drawn, and recommendations are made.

Keywords. Cybersecurity, software, education

1. Introduction to software application development

The development of software systems and platforms is a difficult and lengthy process. This process has a certain number of stages. Each of the individual stages is characterized by a goal, specific activities, indicators of achieved results and duration. The set of stages shows the process of completed software development. It is necessary to pay great importance to the stage of designing the software system, because the technological security requirements should be set there. Training end users to work with the created platform also has its importance. Better trained users can also be more useful in ensuring the level of cybersecurity of the organization. Fig. 1 shows an example of a software development process with relevant stages.



Fig. 1 Software platform development process

The main stages of the software development process described are: Researching the information needs of the organization; Designing the architecture of the software platform; Program implementation; Program testing; Vulnerability testing; User training; Implementation of the software platform. Each of the stages can only be properly implemented if experts with the relevant high-level cybersecurity competencies are provided. To this end, methodologies for assessing competencies can be developed and implemented to serve as a basis for their enhancement [1]. It is extremely important to pay attention to the way the data is organized at each stage. In certain software applications, data storage and retrieval techniques are key to achieving high efficiency of the entire software solution [2]. Documenting each of the stages is of utmost importance, because through it the sustainability of the entire project can be achieved. In each of the stages thus defined, relevant principles and techniques should be established to ensure cybersecurity.

2. Requirements and principles for secure software development

One of the main methods for testing software applications for vulnerabilities is penetration testing. According to some sources, penetration testing is defined as: „Penetration testing (or pen testing) is a security exercise where a cyber-security expert attempts to find and exploit vulnerabilities in a computer system“ [3]. It aims to determine the degree of vulnerability of the software application. Some of the analyzed sources also define the testing process in stages. The phases described are: Pre-engagement, Reconnaissance, Vulnerability Identification, Exploitation, Post-Exploitation, and Reporting [4].

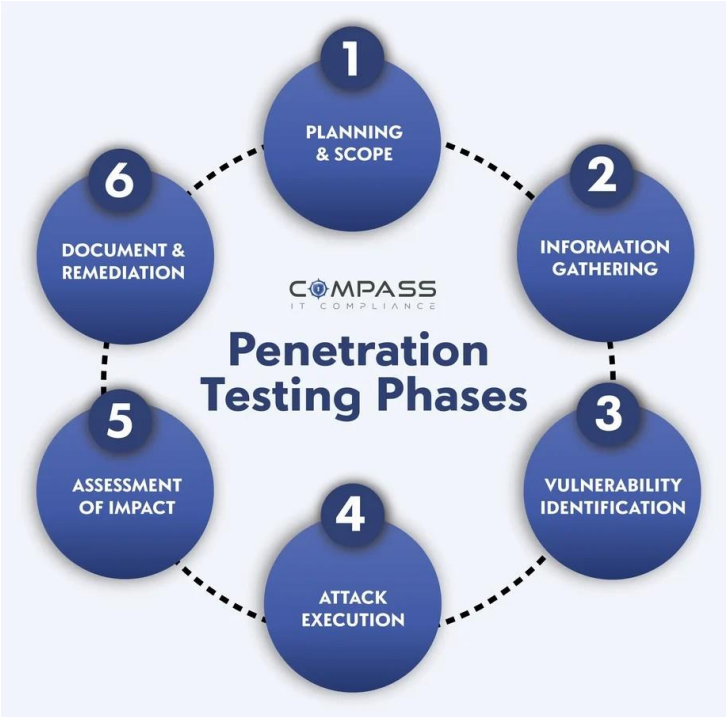


Fig. 2 Penetration Testing Phases [4]

The following important areas of software development should be checked in the individual phases of testing, namely:

- A system for automatic update with new versions of the software has been created. The lack of regular updates, both of the developed system and of the operating system, can be a source of cyberattacks. Therefore, all software components should be updated with each new update.
- The passwords used, and more precisely, the absence of weak passwords. Weak passwords are easy to compromise, which is why they should be avoided or prohibited in the system. It is necessary to emphasize that research shows every year the need to create and maintain increasingly complex and long passwords;
- The prerequisites for social engineering. It is good for the system to be protected from the application of various variants and modifications of social engineering. This is one of the most used methods for penetrating a given software system. It is based on the application of various psychological approaches to deceive the user, whereby he reveals sensitive data or allows unauthorized access to the system. Since the foundations here are psychological, it is necessary to design and implement software measures that allow users to reveal authentication parameters;
- Incorrectly designed, programmed or configured modules of the software platform. It is necessary that the software platform be properly designed and implemented to avoid errors in design or implementation. Incorrect configuration in the system modules can also be a cause of vulnerabilities. Therefore, during the implementation process, special attention should be paid to the individual modules and the data transfer between them;
- Processing of sensitive data. Processing of sensitive data is one of the essential issues for ensuring the cybersecurity of the applications being developed. For this reason, it is necessary to take categorical measures to prevent the unauthorized extraction of sensitive data from the software system.
- Achieving stability in the operation of the system. This is a very important property of the system, which does not allow malicious individuals to penetrate it and maintain control over the functionality. Therefore, at the stage of design and development of the system in its architecture, mechanisms should be created to identify such behavior and at a very early stage to issue the appropriate alarms and notifications.

All this shows that the following important principles should be taken into account and applied even during the development process of software systems (Fig. 3).

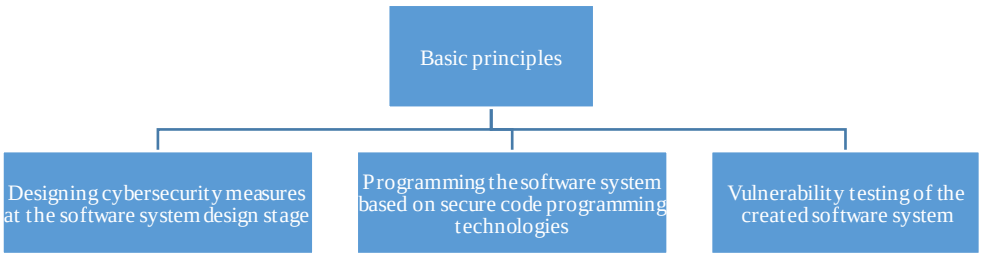


Fig. 3 Basic principles

The basic principles are:

- Designing cybersecurity measures at the software system design stage. It is important to design measures that ensure compliance with cybersecurity rules at the design stage. Adherence to this principle can ensure the architectural construction of a system that adheres to best practices in terms of high access protection;
- Programming the software system based on secure code programming technologies. The programming of the database, business logic modules and user interface elements should be carried out in compliance with the specified cybersecurity measures;
- Vulnerability testing of the created software system.

Compliance with these principles, as well as all the measures and tasks arising from them, when developing software systems can guarantee a high level of security. It is necessary in order to ensure the continuity of the organizations' business processes and achieve high efficiency.

3. Conclusion

Because of the research we can draw the following conclusions:

- In order to ensure a high level of cybersecurity of software systems and applications, it is important that protective mechanisms are built into the architecture at the design stage;
- The programming of software applications must be carried out in compliance with the highest security standards and good practices and ISO 27001;
- Every organization should pay attention to the regular verification of vulnerabilities in software applications and systems, which can guarantee their operation.

References

1. Kirilova, K. (2024). Analysis of Cybersecurity Competencies as A Factor for Organizations' Cyber Defense. In 12th International Scientific Conference on Computer Science (COMSCI), pp. 1-4, IEEE..
2. Milev, P., Guidelines for data model selection and use in web applications for data analysis, Proceedings of International conference on application of information and communication technology and statistics in economy and education (ICAICTSEE – 2018), UNWE - Sofia, 2020, pp. 191 - 194, ISSN 2367-7635.
3. https://www.cloudflare.com/learning/security/glossary/what-is-penetration-testing/
4. https://www.compassitc.com/blog/penetration-testing-phases-steps-in-the-process

Assessment of the Risk Management System of Pension Insurance Companies

Maia Natchkova

University of National and World Economy, Sofia, Bulgaria
mnachkova@unwe.bg

Abstract. The high expectations of society towards the country's social security system determine the level of development and the significance of social security. The main objective of social security is to provide financial security and to maintain the standard of living of insured persons who, due to health-related or other reasons, are unable to work and are unable to receive remuneration from employment during the periods in which they are not engaged in work. Typically, as a result of materialization of the insurable risk and the occurrence of specific insured event, insured persons are unable to work for certain period of time. Insurable risk is an objectively existing, uncertain and unpredictable probability and possibility for occurrence of insured event that causes harm to non-property interest (life, health and working capacity) and results in insured person's inability to earn income from employment while benefiting either from the terms and conditions of the social contract governing the state social security system or of the insurance contract concluded with a private pension fund. The materialization of the insurable risk leads to the occurrence of insured event, regardless of the will and desire of the insured persons. The identification and valuation of the insurable risk for the purposes of determining the amount of the insurance benefit for this risk payable to the insured persons affected by it constitute the principal function of both the state social security system and the private pension funds. The allocation of insurable risk is carried out by the funds of the state social security system and by private pension funds through the organization, management and distribution of dedicated cash funds established from the collected social security contributions paid by employers and insured persons.

Key words: Pension insurance companies, risk management system, insurable risk, insured event, insurance benefit, insured persons.

1. Introduction

Social security is a system of financial protection for employed persons and individuals of non-working age who are unable to support themselves through their own labor upon the occurrence of insurable risk and the resulting consequences, such as incapacity for work due to deteriorated health or childbirth and maternity, or other reasons provided for by the legislator, as well as due to their inability to engage in employment and to meet additional medical treatment expenses.

The high expectations of society towards the country's social security system determine the level of development and the significance of social security. The main objective of social security is to provide financial security and to maintain the standard of living of insured persons who, due to health-related or other reasons, are unable to work and are unable to receive remuneration from employment during the periods in which they are not engaged in work. Typically, as a result of materialization of the insurable risk and the occurrence of specific insured event, insured persons are unable to work for certain period of time.

During this period, insured persons receive social security benefits, allowances and pensions from the accumulated monetary funds administered by the social security institution (the State Social Security system and private pension funds), financed through the social security contributions paid by their employers and by the insured persons themselves.

The identification and valuation of the insurable risk for the purposes of determining the amount of the insurance benefit for this risk payable to the insured persons affected by it constitute the principal function of both the state social security system and the private pension funds. The allocation of insurable risk is carried out by the funds of the state social security system and by private pension funds through the organization, management and distribution of dedicated cash funds established from the collected social security contributions paid by employers and insured persons.

Within private pension funds, insured persons hold individual insurance accounts to which paid compulsory social security contributions are credited on monthly basis. Upon occurrence of insured events, the funds accumulated in these individual accounts are paid in full to the insured person or to their heirs, as they are inheritable. By contrast, under the state social security system, insured persons do not hold individual accounts. Social security contributions are credited into a common money fund (common account), from which cash benefits are disbursed, on the basis of the solidarity principle, to insured persons affected by the occurred insurable risk. The amount of these benefits is determined in advance by the legislature through the annual State Social Security Budget Act for the respective insured events occurred. The common account is not inheritable, however, minor and underage heirs, and adult heirs who are in education are paid survivor's pensions out of the common account, on the basis of the solidarity principle.

The activities of private pension funds are exposed to uncertainty regarding both the timing and the amount of benefits payable to insured persons upon occurrence of insured events under concluded insurance contracts, as well as to market risk related to their pension insurance and investment activities. They do not possess their own capital, personnel and ability to carry out their activities and are therefore managed by pension insurance companies, which are established exclusively for the purpose of managing private pension funds – universal, occupational and voluntary pension funds.

The object of the present study is the issues regarding the assessment of the risk management system of pension insurance companies. The subject of the study is the opportunities for risk management in the activities of pension insurance companies and the activities of the pension funds they manage. The aim is to present the main directions for improvement of the risk management system of pension insurance companies.

To achieve this objective, the author sets the following research tasks:

1) To clarify the essence of the concept of "insurable risk" and to present the main types of insurable risks;

2) To characterize the risk management system of pension insurance companies (PICs) and the methods they use to assess and control insurable risk;

3) To present the main measures adopted by PICs to mitigate and manage risk;

4) To propose directions for improving the risk management system in the activities of PICs

The above tasks may be accomplished by applying various research approaches: historical; logical; system-based; comprehensive, etc.

2. Essence of the concept "insurable risk". Main types of insurable risks

Insurable risk is an objectively existing, uncertain and unpredictable probability and possibility for occurrence of insured event that causes harm to non-property interest (life, health and working capacity) and results in insured person's inability to earn income from employment while benefiting either from the terms and conditions of the social contract governing the state social security system or of the insurance contract concluded with a private pension fund. The materialization of the insurable risk leads to the occurrence of insured event, regardless of the will and desire of the insured persons.

An insurable event is an actually occurring insured event that has already resulted in the inability of the insured person to derive income from their employment. It gives rise to pecuniary consequences for the insured persons, such as loss of income, unforeseen medical treatment expenses and maintenance costs, as well as to personal consequences, including:

- temporary incapacity for work due to deteriorated health (general illness, occupational accident, occupational disease);
- permanent incapacity for work resulting from the natural biological ageing process of the human body (old age and disability);
- need to provide care for a family member (maternity, childcare, care for a sick family member);
- inability to obtain employment (unemployment);
- loss of financial support previously derived from the income of the insured person due to their death. Only in the case of the insurable risk of death do the resulting adverse pecuniary consequences affect not the insured person but their heirs.

Insurable risk arises independently of human will and is random in nature. The only exception to this rule is the risk "Maternity", which arises as a consequence of human choice. Insurable risk manifests itself within society and exists in objective reality, making it foreseeable. It is always social in nature because its consequences are adverse both for the insured persons, who are members of society, and for overall social productivity, as the affected persons are unable to engage in employment for certain period of time due to the occurrence of the insurable event. This risk gives rise to legal consequences under social security law for the social security institution, which becomes obliged to pay social security benefits, allowances and pensions in respect of the social security rights acquired by insured persons upon the occurrence of the insurable event. The social security institution (the funds of the State Social Security system and private pension funds) assumes part of the adverse consequences arising as a result of the occurrence of the insurable risk by paying social security benefits, allowances and pensions to the affected insured persons.

Bulgarian legislation provides for the coverage of all insurable risks. The legislator defines the scope of the State Social Security system, which "provides benefits, allowances and pensions in the event of:

1). temporary incapacity for work;
2). temporary reduced working capacity;
3). disability;
4). maternity;
5). unemployment;
6). old age;
7). death." (Social Security Code, 2025, article 2, paragraph 1)

Insurable risk may be either general or occupational.

General insurable risk may occur to any insured person, irrespective of the type of work they perform. This risk is manifested through the occurrence of the following insurable events: general illness, maternity, old age and death affecting any insured natural person.

Occupational insurable risk arises in the course of dependent employment performed under working conditions provided by another person under employment, civil service or another legal relationship equivalent to employment (e.g. management and control contract). While performing such work, insured persons have no opportunity to influence their working environment. This risk is manifested through the occurrence of the following insurable events: unemployment, occupational accident and occupational disease. The determination of the occurrence of occupational risk is subject to specific procedure and less stringent requirements for acquiring entitlement to social security benefits than those applicable to general insurable risk. For example, in case of the insurable events of occupational accident and occupational disease, no qualifying period of employment is required, as these events may occur even on the insured person's first working day. Occupational risks generally result in more severe consequences for the insured persons than the general insurable risk; therefore, higher amounts of social security benefits are paid in such cases.

Insurable risk may also be classified as inherent risk and residual risk.

Inherent risk is the risk associated with the nature of the activities of the social security institution before any actions (controls) have been implemented to mitigate either the likelihood or the impact of its occurrence.

Residual risk is the level of risk impact and likelihood that remains after the management of the social security institution has implemented its response (controls) to the inherent risk, since risk cannot be eliminated entirely. Consequently, residual risk continues to exist even after specific measures have been taken to prevent its occurrence. It is the responsibility of the management of the social security institution to assess whether the level of residual risk is acceptable or whether new measures or further actions are required to reduce and mitigate it.

Social security benefits are paid from the monetary funds administered by the social security institution (the State Social Security system and private pension funds) upon the occurrence of insurable risk and the corresponding insurable event only where the following conditions are met:

1) The working natural person is compulsorily or voluntarily insured under employment relationship, civil service relationship, service contract or another legal relationship equivalent to employment, or is self-employed person insured against all or certain insurable events;

2) Deterioration of the insured person's health has occurred upon materialization of the risk and the occurrence of the insurable event, i.e. there exists an objective condition in which the integrity of the human body and/or the functioning of its organs has been impaired or damaged. Such impairment is determined by a state medical expert commission. Depending on the manner in which the harm to the insured person's body has occurred as a result of the materialization of the insurable risk, the impairment may be of two types:

a) traumatic injury – the impairment occurs as a single, unexpected, external, unforeseeable event over a short period of time; or

b) disease-related impairment – the impairment develops repeatedly, systematically, permanently, continuously, predictably over prolonged period of time;

3) Existence of incapacity for work – this is established through medical assessment as an objective medical condition by state medical expert commission.

Incapacity for work may be:

a) temporary, in which case the insured person is entitled to receive social security benefits for general illness and maternity; or

b) permanent, in which case the insured person is entitled to pension (disability pension or old-age pension);

4) Existence of direct causal relationship between the impairment and the incapacity for work – this is established through medical assessment by state medical expert commission, and depending on the nature of the causal relationship, the incapacity may result from:

a) general illness; or

b) occupational disease;

5) Existence of relationship between the impairment and the insured person's occupational activity – this is established through medical assessment by state medical expert commission. Such relationship may be:

a) direct causal relationship, in the case of occupational disease; or

b) functional relationship, in the case of occupational accident.

The management of insurable risk includes the identification, assessment, analysis and control of potential adverse events or situations that may occur and negatively affect the life, health and working capacity of insured persons and then the liquidity, solvency and financial obligations of pension funds and the PICs managing them. Risk management should provide reasonable assurance (satisfactory level of assurance) that the risk will be effectively managed, anticipated and controlled, which requires the timely implementation of measures and actions aimed at preventing the materialization of the risk and the occurrence of the adverse event or situation, while ensuring that the costs of internal control do not exceed the expected benefits derived from such controls. Such management may be achieved through a dedicated risk management system established within PICs.

3. Characteristics of the risk management system of pension insurance companies and the methods they use to assess and control insurable risk

PICs are required to design rules for the organization of the activities of each pension fund they manage and for which they have been granted management license by the supervisory authority – the Social Insurance Supervision Division of the Financial Supervision Commission. These include supplementary compulsory pension insurance fund, supplementary voluntary pension insurance fund and supplementary voluntary pension insurance fund operating under occupational schemes. Within these rules, PICs are required to specify the investment policy and investment strategy of each pension fund under their management; the rules governing the payment of social security benefits and pensions to insured persons; the risk management system related both to their own activities and to the activities of the pension funds they manage; the internal control and operational control and internal audit arrangements applicable to the activities of the pension funds they manage.

Private pension funds operate in high-risk environment due to the frequent materialization of insurable, financial and market risks. These risks must be managed through their identification, analysis, assessment and determination of appropriate risk response measures. For this reason, the legislator requires the PICs managing these funds to establish "an effective system of governance that ensures the sound and prudent management of supplementary pension insurance activities. The system of governance shall be proportionate to the nature, scale and complexity of the activities of the pension insurance company and shall include appropriate and transparent organizational structure with clear allocation and proper

segregation of responsibilities, as well as effective system for the transmission of information." (Social Security Code, 2025, article 123e).

"The pension insurance company shall establish a risk management system that covers strategies, processes and reporting procedures for the continuous identification, measurement, monitoring and management of the risks (individually, collectively and in terms of their interdependencies) to which the company and the funds managed by it are or could be exposed." (Ibid., quote reference).

"The risk management system shall, in a manner proportionate to the size and internal organization of the pension insurance company and the funds managed by it, and in accordance with the size, nature, scale and complexity of their activities, cover the risks that may arise within them or the outsourced service providers..., at least in the following areas:

1). provisioning;
2). asset and liability management;
3). investment activities;
4). liquidity risk management;
5). concentration risk management;
6). operational risk management;
7). application of risk mitigation techniques;
8). environmental, social and governance risks related to the investment portfolio of a supplementary voluntary pension insurance fund under occupational schemes and its management, where such risks are provided for in the investment policy of the fund." (Ibid., quote reference)

The risk management system of PICs should be effective and well integrated into its organizational structure and decision-making processes. In addition to insurable risk, it should also address:

• biometric and investment risks from the perspective of the insured persons, pensioners and their heirs participating in the pension funds under their management;
• the risk of disruption of systems, procedures and critical data and functions;
• risks related to investment management and liquidity management;
• risks related to the information system and human resources.

The main criteria that should be observed upon establishing the risk management system are as follows:

• Compliance of POCs' risk management system with the statutory requirements governing the management of pension funds;
• Existence of adequate and transparent organizational structure;
• Clear and appropriate allocation of responsibilities among the governing bodies with respect to the establishment of the risk management system;
• Establishment of effective system for provision of information to stakeholders;
• Ensuring that the risk management system is subject to periodic review, assessment and control;
• Ensuring that POCs' management understands and adheres to the principles of good corporate governance.

PICs' risk management system comprises:

1) Risk management strategy;
2) Assessment of external and internal environment;
3) Systematic risk assessment and monitoring based on territorial distribution of insurable events across individual economic sectors and geographical regions;
4) Maintenance of sufficient liquid assets to cover the respective long-term liabilities of the pension funds under management to insured persons and their heirs;

5) Valuation of the assets and liabilities of PIC and the funds under its management;

6) Performance of actuarial activities;

7) Implementation of measures to mitigate insurable risk aimed at limiting the risk exposures of the pension funds under their management.

The risk management strategy covers the policies, practices, mechanisms and tools through which it is ensured that the risks faced by PIC are appropriately identified and managed. It defines the manner in which PIC addresses risks that pose potential threat to its own activities and to the activities of the pension funds under its management. The strategy determines the methods to be applied in the risk management process; the methods for documenting that process; the responsibilities of the persons involved in the process; the manner in which the risk management activities will be monitored. It is aimed at maintaining and safeguarding the financial stability, liquidity, solvency and profitability of both PICs and the pension funds under its management, thereby ensuring the fulfilment of their obligations towards insured persons and their heirs, as well as towards shareholders, and defining all acceptable categories of risk together with the target levels of acceptable insurable risk.

The assessment of the external and internal environment includes evaluation of the external and internal factors affecting the activities of PIC and creating risks to the achievement of its principal objective – effective management of the pension funds for which it has been granted a license. This assessment is crucial for determining the level of risk to which PIC is exposed.

The factors of the internal environment of PIC comprise its resources, integrated management systems and management mechanisms, such as tangible, human, financial and information resources; its governance and organizational structure; its information system and communication system for interaction with insured persons and regulatory authorities.

The external environmental factors that have the most significant influence on the activities of PIC include:

a) political conditions – political decisions adopted by the legislator that affect the country's social security system, including the enactment of new legislation or the amendment of existing laws and regulations governing social security;

b) economic conditions – the level of economic growth, inflation, household income, labor market development, etc.;

c) demographic factors in the country – mortality, birth rates, migration, etc.;

d) social conditions – social values, national psychology, public attitudes;

e) technology conditions – the level of technological progress, the state of innovation processes, scientific advances and future developments, the degree of development of telecommunications, digitalization of processes;

f) insured persons and their heirs – the customers of the pension funds under management – target groups, attitudes;

g) relations with national and international institutions – cooperation, exchange of information, etc. Among these factors, the most significant ones are the frequent changes in the regulatory framework and the uncertainty created by them.

Systematic risk assessment and its monitoring based on the territorial distribution of insurable events across individual economic sectors and geographical regions is a process involving the determination of the likelihood (possibility) of materialization of the insurable risk and the assessment of the degree of impact (effect) of that risk on the activities of the pension funds and the PICs managing them upon its occurrence. The risk assessment represents the combined result of the assessments of both the impact and the likelihood of each identified risk.

The territorial distribution of insurable events across individual economic sectors and geographical regions highlights "the need for precisely tailored regional strategies that not only take the national context into account as a framework for development, but also fit organically into the unique socio-economic, natural and market environment of each region. Such strategies should be based on the balance between the general priorities of national policy and local needs, making use of flexible mechanisms to respond to changes in the market conditions, demographic trends and infrastructure realities." (Asenov, S., 2026).

The performance of actuarial activities constitutes the most important prerequisite for risk management. Actuarial activities are consistent, continuous and fully integrated into the operations of PIC. Their effectiveness depends on the level of organization, comprehensiveness and thoroughness of the actuarial methods applied in assessing insurable risk, the quality of the procedures performed, the professional competence and commitment of the actuaries responsible for carrying out these activities.

A fundamental prerequisite for effective risk management is the establishment of clear, precise and measurable assessment methods and their complex application.

The methods used by PICs to assess and control insurable risk include actuarial, statistical, mathematical and financial methods, for the purposes of ensuring their long-term solvency and safeguarding the assets of insured persons. These methods are subject to strict regulation by the state supervisory authority. The assessment covers both biometric risks (reduced working capacity, old age and death) and investment risk (negative returns on the financial instruments held). Actuarial modelling is applied by using biometric tables employed in actuarial mathematics. The probabilities of insured persons surviving to specified age are calculated, and their expected remaining life expectancy is estimated. Stress tests and scenarios are used to simulate extreme events. The impact of increasing life expectancy (longevity risk) and of stock market downturns (market risk) on the technical reserves set aside is analyzed. Value at Risk (VaR) is calculated to estimate the maximum expected loss in the portfolio of the pension funds under management over specific period and at given level of statistical confidence.

For the valuation of the assets and liabilities of PIC and the funds managed by it, Asset-Liability Management (ALM) analysis is made by projecting future cash flows arising from social security contributions and pension payments to insured persons. This method ensures that the maturity of invested assets corresponds to the timing of the payments due to insured persons or their heirs.

The legislator requires each PIC to perform its own risk assessment of both the company and the funds under its management at least once a year, and this assessment must comply with the principles of: legality, expediency and effectiveness. Assessment models for qualitative analysis apply to it based on the practical meaning of policies, their understanding, acceptance and adherence as evidence of effectiveness. The assessment includes description of the methods used to identify and assess risk; evaluation of the effectiveness of the risk management system, assessment of the risks to insured persons in relation to the social security payments due to them, updating and recalculation of such payments, assessment of investment risk for investments generating the greatest losses and the adoption of measures to mitigate such risks in the future, qualitative assessment of operational risk and evaluation of the overall funding needs of PIC and the pension funds under its management. The own risk assessment is taken into account by PIC in its strategic decision-making process.

The risk assessment includes the calculation of the Solvency Capital Requirement (SCR) and the Minimum Capital Requirement (MCR) indicators. These two indicators are fundamental for assessing the financial stability and solvency and have been introduced by analogy with the European Solvency II framework applicable to insurance undertakings and

in terms of PICs they ensure that the companies maintain sufficient own funds to absorb unexpected losses and protect the savings of insured persons. PICs are required to exercise continuous monitoring of compliance with the SCR and MCR requirements, as well as with the requirements concerning the amount of the technical reserves maintained. They must also monitor the degree of deviation of the risk profile –deviation of SCR from the Value at Risk (VaR) of their own funds. The control of insurable risk is exercised through the establishment of regulatory mechanisms governing the diversification of investments and the maintenance of mandatory technical reserves.

4. Measures for mitigation and elimination of the adverse effects and for risk management in pension insurance companies

The implementation of measures aimed at mitigating and eliminating adverse the effects and at managing risk in PICs is achieved through the establishment of comprehensive internal control system, diversification of assets and maintenance of mandatory capital reserves in accordance with the requirements of the regulatory authority the Social Insurance Supervision Division of the Financial Supervision Commission (FSC). These measures are intended to protect the long-term pension savings of insured persons against market fluctuations, inflation, operational failures and catastrophic events.

One of the measures for eliminating and mitigating the main risks is the establishment of internal control system. The internal control system of PICs should provide assurance to insured persons and to society, and to the state regulatory authority that supplementary pension insurance activities are carried out in compliance with the legal regulations, the provisions of the internal documents and policies adopted by the company, the concluded insurance contracts and the principles of economy, efficiency and effectiveness of the activities. It should be integrated into the activities of PIC and operate through the governance framework established by its management and supervisory bodies. “The internal control function of the pension insurance company shall be performed by a specialized internal control unit.” (Social Security Code, 2025). “The internal audit function of the pension insurance company shall be objective and independent of the other operational functions, shall be carried out by a separate individual or by a specialized unit, and shall include assessment of the adequacy and effectiveness of the internal control system and of the other components of the system of governance.” (Ibid., quote reference)

Another measure for eliminating and mitigating the principal risks is asset diversification. The allocation of the PICs’ own funds across different classes of asset (shares, bonds, real estate) and across different geographical regions is intended to minimize the losses arising from a single financial instrument. PICs should diversify their investments primarily into instruments with investment credit rating assigned by internationally recognized agencies, low risk and highly liquid assets, and government securities, in order to ensure the availability of cash for the uninterrupted payment of pensions by the pension funds under their management and for the transfer of accumulated funds from the individual accounts of insured persons to other pension funds upon their request. PICs are required to strictly comply with the statutory limits on the maximum percentage of investments in a single issuer or a particular high-risk instrument and to establish credit exposure limits in order to restrict exposures to individual issuers and banking institutions. They should align the maturities of the financial instruments held by matching the maturity structure of their assets with the expected future pension payments of the pension funds under their management to the insured persons. They should also establish a multi-fund system by introducing sub-funds with different investment profiles (dynamic, balanced and conservative), tailored to the age and the risk appetite of insured persons. PICs should conduct regular stress tests by periodically simulating the existence of

extreme market crises in order to assess the resilience of their investment portfolios. They should also employ hedging by using derivatives (futures, options, swaps) in order to ensure protection against sharp fluctuations in exchange rates and interest rates.

The financial risk associated with financial instruments is the possibility of future changes in one or more specified financial indicators, such as interest rate, price of a financial instrument, foreign exchange rate, price or percentage index, credit rating or index, or another variable. This risk may result in adverse consequences and credit losses for PICs. In this regard, D. Feschiyan and R. Andasarova note that "With the adoption of IFRS 9 Financial Instruments..., a new approach to the classification of financial assets was introduced, placing strong emphasis on the contemporary expected credit loss model. The introduced new accounting framework for the recognition of credit risk impairments constitutes a key instrument that has direct impact on the levels of regulatory capital..." (Feschiyan, D., Andasarova, R., 2020, p. 69)

The third measure for eliminating and mitigating the main risks is the establishment of financial buffers through the allocation of specific technical reserves. PICs are required by the state regulatory authority to allocate higher technical reserves to guarantee the gross amount of contributions, thereby ensuring the payment of pensions and covering deficits arising from the management of pension funds. These reserves serve to protect insured persons against market risks and negative investment returns generated by private pension funds, while also protecting PICs against unexpected increase in the life expectancy of pensioners, according to the annual statistics from the mortality table of the National Institute of Statistics. PICs are required to guarantee a minimum rate of return by maintaining special reserve designed to cover potential deviations in the rate of return of the universal and occupational pension funds under their management whenever it falls below the market average.

Other measures that PICs may implement to eliminate and mitigate the main risks include:

- regular identification of new threats arising from the materialization of insurable risk that directly affect PIC and the pension funds under its management;
- annual review of the internal rules and update of the risk identification procedures;
- conducting daily monitoring and analysis of developments in the international financial markets;
- pursuing prudent dividend distribution policy;
- continuous monitoring of the risk profile while taking into account any changes in it, etc.

5. Directions for improving the risk management system in the activities of pension insurance companies

1) Improving the financial management and control systems through digitalization and automation of control, implementation of software for real-time budget monitoring, regular risk assessments, introduction of clear controls and the continuous training of personnel. Responsibilities should be segregated (so that the person performing the verification is not the same person responsible for the execution) and a dual-signature system should be introduced to ensure ex ante control before any financial commitment is undertaken;

2) Enhancing the internal audit by establishing direct reporting line between the internal auditors and the top management of PIC, thereby ensuring the independence of the audit. A risk-based audit should be developed by focusing audit activities on the areas presenting the highest risk of abuse. Software for automated analysis of large volumes of transactions should be implemented. A system for monitoring the implementation of the recommendations contained in the audit reports should also be introduced. Continuous operational monitoring

should be carried out by specialized units within the company to verify that the risk profile remains consistent with both the internal rules and the statutory regulations;

3) Full commitment of the management of PIC to the process of insurable risk management;

4) Maximum integration of data management systems through the use of unified information platforms (for example, the unified information system of the Ministry of Finance), with the aim of improving transparency and disclosure and establishing effective feedback mechanisms that link the findings of the internal audit directly to corrective actions addressing the weaknesses in the risk management system;

5) Minimizing the consequences of the materialization of insurable risk in support of PIC's objective of managing such risk by introducing programs for requalification, for preventive medical examinations, for differentiated medical treatment and medical devices, aimed at reducing unemployment and recovery of the working capacity of insured persons and facilitating their return to the labor market as quickly as possible;

6) Timely identification of manageable and non-manageable insurable risks that may affect the lawful, effective, efficient and sustainable performance of the activities of PIC. This process may be carried out by analyzing the source of risk and determining the degree of control that PIC is able to exercise over it. Non-manageable insurable risks cannot be prevented because they arise from the external environment – from natural or macroeconomic factors. PIC can only compensate for their financial consequences or minimize their impact;

7) Integrated and effective risk management capable of providing useful results for operational management through optimization and introduction of innovative risk management methods, such as the preparation of responsibility assignment matrix (RACI), in which each risk and each response measure is assigned to clearly designated responsible person by name and position. Development of contingency plans for critical crisis situations;

8) Ensuring sufficient advance information concerning the likelihood of the materialization of insurable risk and the occurrence of potential insurable events, the possible methods for mitigating such risk and its potential adverse impact on the activities of PIC;

9) Enhancing the qualification of personnel of PIC, strengthening the knowledge and expanding the competencies of all participants involved in the risk management process through motivation, improved communication and dissemination of good practices;

10) Streamlining the process and updating the risk management rules and methodology through automation and implementation of software tools for automatic data collection and generation of standardized reports (ERM systems). Updating the probability scales for the materialization of insurable risk in accordance with the evolving business reality. Calculation of key risk indicators (KRIs) capable of providing early warning for the threat of materialization of insurable risk before it occurs, rather than reporting only historical events;

11) Promoting the introduction and application of adequate and effective risk control mechanisms. Control mechanisms may be preventive controls, which prevent the occurrence of the risk (e.g. implementation of access control levels), detective controls, which identify the risk that has already materialized and the resulting insurable event (e.g. stocktaking, audit), and corrective controls, which eliminate the consequences of the risk after it has occurred (e.g. data backup). Balance should be maintained in the application of these control mechanisms so that the costs of control do not exceed the potential damages;

12) Optimizing the use of the resources required for insurable risk management by balancing financial expenditures, human capital and technology solutions in order to ensure the long-term financial stability of PIC.

13) Reinsurance aimed at transferring part of the insurable risk to reinsurance companies in order to provide protection against the large-scale materialization of the risk resulting from catastrophic events, etc.

6. Conclusion

The risk management system is a key tool that supports the management of resources and the successful achievement of the objectives of PICs. The fundamental principles upon which the risk management system of PIC should be established are transparency, reliability, comprehensiveness and accessibility of information, coordination and consistency of actions, promotion of communication aimed at providing complete, objective and balanced information, while ensuring cost-effectiveness in order to achieve maximum results. Risk management in PICs should be based on centralized approach, while at the same time this process should be integral part of the overall governance and of its activities related to the management of pension funds.

References

1. Raduilski, L. (1957), Labor and Law of the People's Republic of Bulgaria, "Nauka i Izkustvo", Sofia, 1957, p. 677.
2. Social Security Code, (2025), promulgated in SG no. 110/17.12.1999, effective from 01.01.2000, last amendment SG no.100/25.11.2025.
3. Accounting Act, (2025), promulgated in SG no 5 of 08.12.2015, effective from 01.01.2016, last amendment SG no 115/30.12.2025.
4. International Financial Reporting Standards (IFRS), (2022), Regulation (EC) No 1126/2008 of the European Parliament and of the Council of 3 November 2008 adopting certain international accounting standards in accordance with Regulation (EC) No 1606/2002 of the European Parliament and of the Council
5. Natchkova, M., (2022), Specific disclosures in the financial statements of the pension insurance companies in conditions of crisis, International journal „KNOWLEDGE", Institute of Knowledge Management, Skopje, Scientific Papers, Vol.50.1, 02/2022, p. 45-51, ISSN 2545-4439, ISSN 1857-923X
6. Asenov, S. (2026). Influence of Territorial Factors and Dependencies on the Dynamics of Agricultural Land Prices in Bulgaria. Bulgarian Journal of Agricultural Economics and Management, 71(1), 3-22 (BG). https://doi.org/10.61308/FZQR5658
7. Feschian, D., Andasarova, R., (2020), „Quality of Assets and Regulatory Capital – the New Challenge faced by the Bank Management", Collection of Reports from the Anniversary International Scientific Conference on the occasion of the 100th Anniversary of the Accounting and Analysis Department: "Contributions of Accounting to the Economic Science", Publishing House UNWE, S., 2020, p. 69.

Using Mathematical Methods to Solve Economic Problems and Production Crises in Energy and Tourism

Kostadin Sheiretsky, Svetlin Antonov

Technical University of Sofia
sheiretsky@abv.bg

Abstract. The article examines the main factors of development in the energy sector and tourism. It provides an analysis of the structure of these systems and discusses opportunities to apply optimization in decision-making to address problems that arise in both economic sectors. Examples of simple mathematical models are given.

Keywords. Mathematical methods, Economic problems, Energy, Transport, Tourism

1. Introduction

Modern times require making political and economic decisions related to adapting to changing natural and social conditions. We face challenges posed by a changing climate - droughts, rising temperatures, and rising sea levels. Geophysical changes lead to extreme atmospheric phenomena and changes in oceanic and sea circulation. These factors directly affect aviation, agriculture, and tourism and indirectly influence nearly all other industries.

The growing global population results in food shortages, a lack of drinking water sources, and migration problems. Increasing inequalities in the economic potential of different countries often trigger political and military conflicts. Technological development also poses new challenges for the economic system. Machines increasingly replace people, while new professions arise, requiring specialized worker training.

The above issues are only part of what necessitates real changes in the economic and political systems of civilization. Recent events show that the adaptability of today's economic system is diminishing. There is talk of an evolution that this system must undergo in accordance with the new conditions. In this context, economic decision-making cannot rely solely on political arguments [1].

Moreover, maintaining the sustainability of the social world we live in requires optimization and the introduction of new technologies in key economic sectors. Here lies the role of mathematics - as a fundamental tool for management and predictive modeling. Based on mathematical relationships, such models provide information about the effectiveness of decisions made in a changing environment. This paper discusses several key sectors where changes must be made according to mathematically derived conclusions [2].

2. Energy Sector

Over time, the population's energy needs have increased. Recently, the use of renewable natural resources has been promoted, but current technology levels do not yet make them sufficiently efficient. Oil remains dominant. The use of nuclear energy is more interesting. Modern reactors generate energy from the fission of heavy nuclei. The next stage is to use the binding energy of light nuclei—so-called thermonuclear fusion. Unfortunately, such reactors are not yet stable or economically efficient enough for practical use [3, 4].

What is the role of mathematics in solving energy-related problems?

- Finding an optimal balance between different types of energy sources from an economic point of view;
- Assessing the ecological value of various energy-production technologies;
- Making decisions aligned with safety requirements for different types of power plants.

Nuclear power plants meet nearly all criteria for a modern energy source: high efficiency and environmental friendliness when properly operated. A technological drawback is that their byproducts are highly radioactive and must be stored under special conditions. The main problem in nuclear energy is the catastrophic consequences of potential accidents. History shows this can result from human error, as in Chernobyl, or natural disasters, as in Fukushima.

The above are only part of the criteria for decision-making in strategic energy planning. As members of the European Union, such decisions are far more effective when made globally for the entire continent. Mathematically, this can be achieved through multicriteria optimization models, some of whose parameters have a probabilistic nature [5]. To this end, several steps are necessary:

- Development of a strategic plan for the EU's energy evolution within reasonable timeframes;
- Research on the environmental efficiency of existing energy technologies. For example, wind turbines can alter local air circulation and potentially cause ecological crises in certain regions, despite their advantages. Decisions must therefore be made regarding their number and density. Tidal power plants can change current dynamics and even lead to serious climatic changes. Large solar panel fields can cause similar disturbances.
- Development of new technologies that improve the safety and efficiency of energy production.
- Development of methods to reduce the energy consumption of electrical devices.

Optimization aims to maximize a given objective function. However, in energy, it is difficult to prioritize one criterion over another since all are crucial. Considering the probabilistic nature of many parameters -and, to a great extent, the dominance of political approaches - the task becomes extremely complex. Even if this were the only issue, it would still be largely solvable. The deeper problem lies in science's incomplete understanding of natural processes. The most vivid example is the ongoing debate: what actually causes global warming - anthropogenic changes or planetary processes beyond our influence?

3. Tourism Sector

Bulgaria's geographic location enables both mountain and seaside tourism [6]. The problems in tourism can be divided into several groups:

Environmental issues: overdevelopment of coastal areas, discharge of wastewater into the sea, pollution from rivers and vessels, etc. In recent years, mountain regions have faced numerous forest fires, which cause not only catastrophic events but also large-scale air pollution. Floods frequently damage infrastructure and endanger human lives.

Service quality and offered products: we include these due to the complexity of providing effective tourism services.

Transport system: this is of critical importance to tourism. Practically, investment in railways and road construction is an essential investment in tourism itself.

What is the role of mathematics here? Primarily, optimization of state investments in areas of tourist importance. Funds must be allocated to effectively solve the above problems. Tourism plays a vital role in stimulating the development of many regional industries—typically businesses related to services and trade. Many of these are strongly affected by the seasonal nature of regional tourism.

A key task is developing a strategy that defines how investment should be made. Political factors often complicate decision-making regarding feasibility. Although politics can be incorporated into mathematical models, the issue lies in the abrupt and often unjustified changes in political decisions, turning them into probabilistic factors. This conclusion does not diminish the importance of politics in economics. On the contrary, they are inseparable. With properly planned policies and a stable government, certain regions can experience rapid economic growth.

One of the main reasons for the economic underdevelopment of northwestern Bulgaria is the weak tourism in the region. It is therefore worth assessing what kinds of investments would be appropriate to stimulate businesses and improve economic well-being - such as increasing transport flows through new bridges over the Danube River. Naturally, such decisions must be taken while the region still retains adequate infrastructure and human resources, considering its rapid depopulation [3].

4. Elementary Mathematical Models with Applications in Transport and Energy Economics

One of the most widely used models describing the adaptation of two interacting entities is the Lotka–Volterra model. Although it is best known as the predator – prey model, it is isomorphic not only to biological systems but also to economic, political, and social systems. In energy and transportation systems, this model can be successfully applied to clarify the dynamics and interactions between two entities.

By introducing nonlinear terms into the governing equations, it becomes possible to investigate not only steady-state processes but also bifurcation phenomena. Owing to the model's fundamental importance, a brief overview is presented below [7, 8].

As a specific example, let us consider the nonlinear model proposed by Goodwin (1967) [7], also known as the *class struggle* model, which belongs to the family of Lotka-Volterra systems.

Let V denote the number of workers in the transportation sector, and let U enote the number of employers. The model is described by the following system of equations:

$$\dot{U} = (-a + b.V)U \tag{1}$$

$$\dot{V} = (c - d.U)V$$

where the constants a , b , c , and d are positive, and the dot denotes differentiation with respect to time.

This model exhibits persistent oscillatory behavior. Its structural instability can be eliminated by adding to the expression in parentheses in the first equation of system (1) a function of U whose first derivative is negative over the domain of variation of U .

One of the earliest nonlinear models in economics was proposed by Kaldor [7,8]:

$$\dot{Y} = \alpha[I(Y,K) - S(Y,\ K)] \tag{2}$$

$$\dot{K} = I(Y,K) - \delta.K$$

where K denotes the fixed capital in a given sector of the economy, Y represents income, I denotes investment, S represents savings, and α and δ are constants.

The dot denotes differentiation with respect to time. It is assumed that:

$$\frac{\partial I}{\partial Y} > 0\ ,\ \ \frac{\partial I}{\partial K} < 0\ ,\ \ \frac{\partial^2 I}{\partial Y^2} > 0\ (< 0)\ ,\ \ I < Y_1\ (> Y_1)\ . \tag{3}$$

Under these assumptions, system (2) may exhibit a self-sustained oscillatory (limit-cycle) solution.

The Kaldor model has also been developed in a discrete-time formulation.

$$Y_{n+1} = \alpha\left[I\left(Y_n, K_n\right) - S\left(Y_n, K_n\right)\right] + Y_n\ , K_{n+1} = I\left(Y_n, K_n\right) + \left(1-\delta\right)K_n\ . \tag{4}$$

The system of equations (4) is capable of generating chaotic behavior.

In [9], a generalized model for the mathematical representation of the aggregate social product is proposed. The following differential equation is introduced:

$$m(x,t)\frac{d^2x}{dx^2} + B\frac{dx}{dt} + C(x) = F(t), \tag{5}$$

where x denotes the aggregate social product, $F(t)$ is the total volume of investment, $C(x)$ is the capital accumulation function, $m(x,t)$) is a measure of the system's inertia, and B is a coefficient that depends on the system's ability to resist changes induced by investment.

To simplify the model, the following assumptions are made:

$$m = const, C(x) = cx + \gamma x^3,$$

where c and γ are constants. Then equation (5) reduces to:

$$m\frac{d^2x}{dx^2} + B\frac{dx}{dt} + cx + \gamma x^3 = F(t). \tag{5}$$

If we assume $F(t) = A\sin\omega t$, the Duffing equation is obtained.

Mathematical models are particularly valuable because they not only describe a process but also make it possible to identify the conditions under which a crisis or a system's evolution may occur. Such models are mainly nonlinear and require specialized methods of analysis. Analytical solutions are difficult to obtain; they are often derived through asymptotic expansions, while in most cases the analysis is carried out numerically.

It is very important that the model parameters are chosen so that the model best matches the measured values of quantities evolving over time. In this way, it is often possible to make assumptions or predictions about the behavior of the system under certain unfavorable conditions. Thus, control can be achieved if the system is controllable.

5. Conclusion

The analysis shows that solving problems arising in energy and transport requires long-term planning. It is essential to invest in applied scientific research, as the solutions to many crises affecting both sectors lie in introducing new technologies. Political decisions should align with optimization principles provided by mathematical theory.

The conclusion is that the management of the energy and transport industries should be conducted by interdisciplinary specialists - economists, physicists, mathematicians, ecologists, etc. For adequate decision-making, mathematical justification and competent program design by interdisciplinary teams are necessary.

Acknowledgments

This work has been carried out in the framework of the Project "Application of Artificial Intelligence for Sustainable Business Development in the Context of Climate Change", financially supported by the Science Fund of the University of National and World Economy, Bulgaria (Grant No. NID NI 35/2025/A).

References

1 C. Chiang and K. Wainwright, Fundamental Methods of Mathematical Economics, 5th ed. New York, NY, USA: McGraw-Hill Education, 2021.
2 D. P. Bertsekas, Nonlinear Programming, 3rd ed. Belmont, MA, USA: Athena Scientific, 2016.
3 R. L. Keeney and H. Raiffa, Decisions with Multiple Objectives: Preferences and Value Tradeoffs. Cambridge, U.K.: Cambridge Univ. Press, 1993.
4 J. Conejo, M. Carrion, and J. M. Morales, Decision Making under Uncertainty in Electricity Markets. New York, NY, USA: Springer, 2010.
5 Jebaraj and S. Iniyan, "A review of energy models," Renewable and Sustainable Energy Reviews, vol. 10, no. 4, pp. 281–311, 2006.
6 N. Vanhove, The Economics of Tourism Destinations, 3rd ed. London, U.K.: Routledge, 2017.
7 Panchev, St. Chaos Theory. Sofia: Prof. Marin Drinov Academic Publishing House, 1996. (in Bulgarian)
8 Zhang, . Synergetic Economics: Time and Change in Nonlinear Economic Theory. Moscow. Mir, 1999) (in Russian)
9 Petrov, L. (2012). Methods of Dynamic Analysis of Economics. Moscow: Infra-M.

Wildfires as Economic Disruptors: Sustainable Policies for Energy, Tourism and Regional Recovery

Kostadin Sheiretsky, Svetlin Antonov

Technical University of Sofia
sheiretsky@abv.bg

Abstract. Wildfires have become one of the most serious environmental and economic challenges of the 21st century, disrupting national economies, destroying critical infrastructure, and threatening key sectors such as energy and tourism. This paper examines the multifaceted economic consequences of large-scale wildfires and analyzes the policies necessary to build regional resilience and foster sustainable recovery. The study focuses on two primary dimensions: first, the direct and indirect economic losses resulting from wildfire disasters and second, the policy, institutional, and technological measures that can mitigate these impacts. Through a qualitative analysis of European and global case studies, the paper identifies how integrated disaster management, renewable energy diversification, and sustainable tourism planning can minimize vulnerability and support long-term economic stability. It also explores the critical role of multi-level governance, public-private partnerships, and community-based initiatives in rebuilding affected regions. The findings highlight that sustainable economic resilience to wildfires depends not only on immediate response capacity but also on proactive preventive investments, risk-awareness education, and strategic regional planning aligned with the principles of the European Green Deal and global climate adaptation strategies.

Keywords: Wildfires, Economic resilience, Sustainable policy, Energy, Tourism, Disaster recovery, Regional development

1. Introduction

The increasing frequency and intensity of wildfires worldwide represent a formidable challenge for governments, industries, and communities alike. This phenomenon, exacerbated by climate change, is no longer a seasonal or regional anomaly but a persistent global threat.

In Europe, the past decade has seen a dramatic escalation in wildfire-affected areas, particularly across Mediterranean countries such as Greece, Spain, and Portugal, which have suffered devastating ecological and material losses. These events cause significant environmental damage and trigger far-reaching socio-economic consequences that disrupt essential sectors including energy, transport, agriculture, and tourism.

In this context, understanding the complex economic impact of wildfires and developing robust, sustainable policies for prevention and recovery is of paramount importance.

This paper aims to analyze wildfires as potent economic disruptors and to propose a framework of sustainable policies designed to enhance the resilience of the vital energy and tourism sectors. By applying a qualitative approach that reviews exemplary case studies and foundational EU policy frameworks, this study synthesizes knowledge on disaster management and sustainable development to offer actionable insights for policymakers and regional planners.

2. The Economic Impact of Wildfires

The economic consequences of wildfires are profound and extend well beyond the immediate, visible destruction of infrastructure and ecosystems. These impacts can be broadly categorized into direct, indirect, and intangible costs, which together create a substantial financial burden on society.

Direct Costs

Direct costs are the most immediately quantifiable and include the expenses of fire suppression, the destruction of residential and commercial properties, damage to public infrastructure like roads and utilities, and the loss of agricultural and timber resources. The World Bank estimates the annual global cost of wildfires now exceeds $50 billion USD, a figure encompassing property destruction, reduced agricultural output, and public health expenditures related to fire impacts.

Indirect Costs

However, the indirect impacts often exceed these direct damages in the long term. These include the disruption of critical supply chains, loss of income for businesses forced to close, and significant declines in tourism revenue. For instance, recent analysis shows that major wildfires in Southern Europe have been responsible for reducing regional GDP growth rates by as much as 1.5% in affected years, demonstrating a tangible brake on economic progress.

Furthermore, the immense fiscal burden of post-disaster recovery and reconstruction places long-term pressure on national and regional budgets, diverting funds from other essential public services. A concerning trend highlighted by insurance data is the significant gap between insured and uninsured losses, which indicates that existing financial protection mechanisms remain critically underdeveloped, leaving many communities and small businesses financially exposed.

3. Implications for the Energy Sector

The energy sector is uniquely vulnerable to wildfire events. Modern energy systems face direct threats from fires that can damage or destroy physical assets such as transmission lines, electrical substations, and renewable energy installations.

In recent years, countries like Spain and Italy have reported multiple instances of solar farms and wind turbine parks being temporarily shut down or damaged due to advancing fire fronts, interrupting energy supply and causing revenue loss.

This risk to grid stability underscores a critical need to transition towards more resilient energy infrastructure. Key strategies include the development of decentralized energy models and microgrids, which are capable of operating independently ("islanding") during a crisis, thereby ensuring continuous power for critical facilities like hospitals and emergency response centers.

Moreover, enhancing economic diversification within the energy sector is crucial. Strategic investments in a mix of renewables, emerging hydrogen technologies, and localized energy storage solutions can significantly enhance systemic resilience and reduce over-dependence on centralized, vulnerable infrastructure. This approach aligns directly with European policy frameworks like the REPowerEU Plan, which explicitly calls for accelerating the transition towards secure, sustainable, and climate-resilient energy systems across the continent.

4. Challenges and Strategies for the Tourism Sector

Tourism is one of the economic sectors most immediately and severely affected by wildfires. The combination of environmental degradation, hazardous air pollution from smoke, and acute safety risks leads to mass cancellations and a sharp decrease in visitor numbers.

The reputational damage to a destination can be profound and long-lasting. According to the UN World Tourism Organization, destinations impacted by major wildfires can lose up to 40% of their tourist inflows during the subsequent season, a staggering loss for regions often heavily dependent on tourism revenue.

A successful recovery requires a coordinated and multi-faceted strategy aimed at restoring public confidence, rebuilding essential infrastructure, and actively promoting sustainable tourism practices.

This is not merely about reconstruction but about reimagining a more resilient tourism model. Strategic investments in eco-tourism, biodiversity conservation projects, and the development of certified safety zones are essential components of this approach, helping to attract environmentally conscious travelers and demonstrate a commitment to safety and sustainability.

The revival of affected areas hinges on strong regional cooperation among municipalities, tourism boards, and private investors, whose collaborative efforts are crucial for ensuring the sector's long-term viability.

5. Integrated Policy and Governance Strategies

Effective wildfire management is impossible without integrated policy frameworks that cohesively link environmental protection with economic resilience. This requires a multi-level governance approach, from transnational bodies to local communities.

At the EU level, instruments like the Civil Protection Mechanism and the European Climate Law (2021) provide a strong foundation, offering financial resources and institutional support for member states to enhance their preparedness and response capabilities. Flowing from this, national governments are encouraged to adopt proactive and preventive strategies that combine modern early warning systems, sustainable land management practices (such as prescribed burns and forest thinning), and sophisticated risk mapping to identify vulnerable areas.

Mobilizing the necessary capital for these initiatives often requires innovative funding models. Public-private partnerships (PPPs) can leverage private sector investment and expertise for large-scale prevention and recovery projects, from building fire-resilient infrastructure to developing new firefighting technologies.

Finally, the importance of grassroots action cannot be overstated. Community-based programs, which empower local residents with knowledge and tools, have proven highly effective in raising risk awareness and improving local preparedness, creating a first line of defense against wildfire threats.

6. A Framework for Sustainable Recovery and Long-Term Resilience

Moving beyond immediate response, building long-term economic resilience to wildfires requires a paradigm shift: integrating disaster management into the core of broader sustainable development policies. This "building back better" approach focuses on creating stronger, more adaptable communities and economies.

A range of economic instruments can be deployed to support this vision. Well-designed insurance products can cushion financial losses, while tax incentives can encourage businesses and homeowners to invest in green, fire-resistant infrastructure. In regions where tourism and agriculture are the primary economic drivers, diversification and innovation become essential for survival.

This includes promoting agritourism based on fire-resilient crops or developing digital nomad hubs to attract a less seasonal workforce. Investments in digital infrastructure, decentralized renewable energy, and vocational training programs are critical for reducing economic vulnerability and promoting inclusive, sustainable growth.

By explicitly linking wildfire recovery efforts with the ambitious goals of the European Green Deal, nations can seize the opportunity to accelerate their green transformation, enhance regional cohesion, and build a more resilient future.

7. Conclusion

Wildfires pose not only a severe environmental danger but also a profound economic threat to regional and national stability. A reactive, crisis-driven approach is no longer sufficient or economically viable.

The analysis presented in this paper demonstrates that achieving sustainable resilience requires a deeply integrated strategy that combines multi-level policy coordination, strategic economic diversification, and robust community engagement.

The energy and tourism sectors, being both highly vulnerable and critical to regional economies, must be prioritized in all prevention and recovery planning. By systematically aligning disaster management measures with the principles of sustainable growth, European economies can transform the challenge of wildfires into a powerful catalyst for innovation, resilience, and lasting social progress.

Acknowledgments

This work has been carried out in the framework of the Project “Smart Risk Management for Business from Adverse Events and Natural Disasters”, financially supported by the National Science Fund of Bulgaria, Grand № KP-06-H55/5 from 15.11.2021.

References

1 European Commission, REPowerEU Plan, Brussels: Publications Office of the European Union, 2023.
2 European Environment Agency, Wildfire Risk in Europe: 2023 Assessment, Copenhagen: Food and Agriculture Organization of the United Nations (FAO), *Community-Based Disaster
3 International Energy Agency (IEA), Energy Resilience and Wildfire Risk, Paris: OECD/IEA
4 Munich Re, Natural Catastrophe Report 2024, Munich: Münchener Rückversicherungs-Gesellschaft, 2024.
5 Organisation for Economic Co-operation and Development (OECD), Climate Risks and Economic Resilience in Europe, Paris: OECD Publishing, 2023.
6 United Nations World Tourism Organization (UNWTO), Tourism and Climate Resilience, Madrid: UNWTO Publications, 2023.
7 World Bank, The Economic Cost of Wildfires, Washington, DC: World Bank Group, 2022.

Data as a Governance Tool: Big Data and SME Export Promotion in South Korea

Hannah Kang

University of National and World Economy, Sofia, Bulgaria
hannah.kang@unwe.bg

Abstract. While some accounts suggest that the developmental state has lost relevance, South Korea's use of digital and data-based instruments in export promotion shows how the state is adapting its developmental role to the conditions of the digital economy. This paper asks how big data has been institutionalised as an instrument of export promotion in Korea, and what the process reveals about the adaptation of developmental governance. Korea is examined as a paradigmatic case through documentary analysis of Korean- and English-language policy documents, platform materials and official statistics. The analysis traces the convergence of the post-2020 national data strategy with a six-decade-old export promotion apparatus, examining in depth KOTRA's TriBIG platform, which algorithmically matches SMEs with foreign buyers. The findings indicate continuity of developmental ambition pursued through new instruments and delivered through institutionalised public-private data coordination. However, SME adoption of these platforms remains uneven. Because algorithmic tools require analytical capabilities that firms possess to varying degrees, the primary constraint is not the technology itself, but the challenge of aligning technological provision with user capacity. The paper's central contribution is to reframe data not merely as an object of regulation, but as an active instrument of economic governance that the state mobilizes to achieve its developmental objectives.

Keywords. Developmental state; big data; export promotion; SMEs; South Korea; data governance

1. Introduction

The article employs a qualitative single-case design, treating Korea as a paradigmatic case: it combines the archetypal developmental state inheritance with digital government capacity the OECD ranks among the most advanced of its members (OECD 2025). The analysis rests on documentary evidence - government policy documents and legislation, the trade promotion agencies' platform documentation and reports, and official trade statistics - with Korean-language sources permitting triangulation beyond the English-language material on which most treatments rely. The object is institutional: how data-driven export promotion is designed, governed and delivered, not its causal impact on export performance, which the documentary record cannot identify.

The East Asian developmental state emerged in a historical context defined by industrial upgrading, export discipline, and close coordination between the state and large business groups. Its classical instruments - directed credit, trade protection, and performance-based support - were designed to accelerate the movement of selected firms into higher value-added sectors. In the Republic of Korea, or South Korea, these instruments were deployed with particular effectiveness: the *chaebol* became globally competitive exporters, while the state's capacity to discipline and coordinate them became central to the developmental state literature.

This presents the state with a puzzle its classical toolkit cannot solve. What excludes SMEs from world markets is not only capital but information: the fixed costs of identifying promising markets, locating credible buyers and deciphering foreign standards weigh disproportionately on firms too small to internalise them. Subsidies and credit, the instruments that built the *chaebol*, do not remove these informational barriers. A firm that cannot find its buyer cannot usefully be financed to supply one.

This article argues that Korea's response amounts to the emergence of a *digital developmental state*, understood here as the conjunction of three elements: enduring developmental ambition, the deployment of data as a policy instrument in its service, and delivery through embedded public-private coordination. The third element distinguishes the phenomenon from competent e-government of the kind found in many advanced administrations. Through the Korea Trade-Investment Promotion Agency (KOTRA), Korea's state trade and investment promotion agency, six decades of trade statistics, overseas office reporting and client records are being converted into platforms that recommend markets and match buyers, supplying market intelligence as a public good and targeting the information asymmetries that exclude SME exporters.

The article accordingly asks how big data has been institutionalised as an export promotion instrument, and what this reveals about the adaptation of developmental governance. Its answer is developed through three propositions: that the platforms represent continuity of developmental ambition through new instruments; that they operate through public-private data coordination; and that their reach is decided by SME uptake, so that coordination rather than technology constitutes the binding constraint.

Korea is examined as a paradigmatic case, combining the archetypal developmental state inheritance with digital government capacity of the first rank (OECD 2026), and the argument addresses a gap in the literature, which has theorised data as an object of governance far more thoroughly than as an instrument of it. While KOTRA anchors the analysis, the paper examines it as the operating core of a broader data-driven trade support ecosystem. Specifically, it investigates how KOTRA has converted its accumulated informational assets into an algorithmic matchmaking service for SME exporters, examining the platform's design, coordination mechanisms, and patterns of SME uptake.

2. Theoretical Framework: the developmental state in the data economy

This paper begins from the renewed scholarly and policy debate concerning the state's role in guiding techno-industrial development. For several decades, industrial policy remained marginal to mainstream economic discourse, particularly in Western academic circles and international financial institutions. However, since the global financial crisis of 2008, and accelerating further after the COVID-19 pandemic, industrial policy has returned decisively to the centre of both scholarly inquiry and governmental strategy. This resurgence has generated substantial revisionist literature in comparative and international political economy (Bulfone 2023), alongside a parallel rehabilitation of industrial policy within the economics profession itself (Juhász, Lane and Rodrik 2024). Governments across advanced and emerging economies alike are now actively deploying subsidies, local content requirements, and strategic investment programmes to shape industrial outcomes, marking a decisive break with the neoliberal orthodoxy that dominated the late twentieth century.

Within this broader revival of state activism, the developmental state tradition occupies an ambiguous and contested position. On one hand, the classical East Asian developmental states - Japan, South Korea, Taiwan, and Singapore - achieved historic economic

transformations in the post-war period through precisely the kinds of proactive industrial policies now being reconsidered. On the other hand, contemporary scholarship reports a puzzling contradiction: the developmental state appears to be simultaneously disappearing in East Asia, its supposed heartland, while proliferating across other regions, from Latin America to Africa and Southeast Asia. Weiss and Thurbon (2021) attribute this conceptual confusion to the developmental state having become unmoored from its theoretical and historical grounding. Scholars and policymakers, they argue, have conflated the *essence* of developmentalism - an enduring ambition to promote national techno-industrial advancement - with its historically specific *instruments*, such as directed credit, five-year plans, or protected domestic markets. When these particular tools fell out of favour or were retired, many observers prematurely declared the developmental state itself obsolete, failing to recognise that the underlying ambition might persist even as the means of pursuing it evolved. The foundational statements are Johnson (MITI and the Japanese Miracle, 1982) on plan-rational bureaucracy and the pilot agency; Amsden (Asia's Next Giant, 1989) on disciplined reciprocity between the Korean state and business; and Evans (Embedded Autonomy, 1995), whose account of dense yet autonomous state–business ties remains the touchstone for the public–private coordination examined below.

In response to this conceptual impasse, Weiss and Thurbon (2021) propose a significant analytical shift. Rather than searching for fixed institutional forms - the presence of a pilot agency, indicative planning mechanisms, or specific financial controls - scholars should redirect their attention towards what they term *developmental governance*: purposive state activism at the techno-industrial frontier, shaped by national ambition and mediated through the specific configuration of state–society relations in which it is exercised. This reframing emphasises *what states are trying to achieve* and *how they organise themselves to pursue those goals*, rather than whether they possess a predetermined checklist of institutional features inherited from the East Asian miracle era. It is this refurbished conception, which prioritises strategic intent and adaptive capacity over rigid institutional typologies, that frames the present analysis. By adopting this approach, the paper can trace how developmental ambition persists even as the instruments of its pursuit transform in response to new technological and economic realities.

Two implications of this governance-focused approach matter here. The first concerns *continuity* of the developmental ambition. Thurbon's (2016) account of the developmental mindset locates the essence of developmentalism not in any given instrument - whether directed credit, trade protection or five-year economic plans - but in an enduring official ambition to promote national techno-industrial advancement. This ambition survives the retirement of old instruments and animates the design of new ones. The second implication, which follows directly from the first, is the capacity for *domain migration*. Because the developmental mindset is an enduring ambition rather than a fixed historical toolkit, it can be redirected toward entirely new policy frontiers. Thurbon, Kim, Tan and Mathews (2023) demonstrated this by showing how East Asian states applied their classical developmental drive to the green energy transition - a dynamic they term developmental environmentalism. Their work provides the analytical template for this paper: just as the developmental state migrated into environmental policy to secure national advantage, this paper traces how it is now migrating into the data economy, applying the same enduring ambition to a new class of digital instruments.

A second body of recent scholarship examines the state's encounter with the data economy itself. At the global level, data has been theorised as a contested object of regulatory competition between rival governance models (Bradford 2023) and as a development resource whose benefits are unevenly distributed across and within countries (UNCTAD 2024). Korea

figures prominently in this literature as a limiting case of state digital capacity: the OECD's Digital Government Review of Korea (2025) documents how the Korean public sector has embedded digital technologies and data in governance and service delivery for decades, sustained by successive national strategies and statutory five-year master plans. Most directly relevant, Eom, Ko, Park and Park (2025) analyse what happens when the Korean developmental state meets the data economy, showing a state that balances regulatory oversight with market facilitation in pursuit of data-driven growth while contending with fragmented data-sharing mechanisms and pronounced disparities in analytical capability across enterprises.

Reading these literatures together exposes the gap this paper addresses. Much of the literature in international political economy has focused on data as an *object of governance* - something to be regulated, localised, traded or protected - leaving its role as an *instrument of governance* underexplored: a resource the state itself mobilises to perform its economic functions. Where the digital-era developmental state has been examined (Eom et al. 2025), the analysis concerns the data economy as a sector to be cultivated. The deployment of big data as a policy instrument within a classical developmental function, export promotion, remains unexamined. That is the space this paper occupies.

The economic rationale connecting data to export promotion is supplied by the literature on information frictions in trade. Rauch (1999) demonstrates that information deficits constitute a fundamental barrier to international trade, with business and migrant networks serving to reduce these frictions. Exporting entails substantial fixed informational costs, from identifying promising markets and locating creditworthy buyers to deciphering foreign standards and procedures, and these costs weigh disproportionately on small and medium-sized enterprises, whose scale precludes internalising them. Export promotion agencies exist to help firms share these costs. While they successfully boost exports, their impact is strongest at the initial stages of market entry, which shows that their main role is providing information rather than funding (Lederman, Olarreaga and Payton 2010). Big-data platforms intensify this informational support for market entry. Exporting entails substantial fixed informational costs, from identifying promising markets and locating creditworthy buyers to deciphering foreign standards and procedures. Big-data platforms address these frictions by deploying algorithms to match firms with foreign buyers and predictive models to estimate the probability of successful transactions. Yet such instruments presuppose a degree of absorptive capacity on the part of their beneficiaries: recent evidence indicates that a majority of Korean manufacturing SMEs remain unaware of the significance of digital transformation for their operations (Chun 2024), foreshadowing the participation problem examined in the discussion.

From this synthesis the paper derives three propositions. First, *continuity of ambition*: Korea's big-data trade instruments represent the contemporary expression of developmental governance rather than a departure from it. Second, *coordination as delivery mechanism*: the instruments operate through public-private data flows and institutional cooperation, reproducing in informational form the embedded state–business relations that the classical literature identified as the developmental state's distinctive asset. Third, *participation as the test*: the reach of digital developmental governance is measured not by the sophistication of its platforms but by the capacity of SMEs to use them.

The material in Sections 4 and 5 bears principally on the first two propositions; the third is assessed in Section 6.

3. Methodology

The article employs a qualitative single-case design. Korea is treated as a paradigmatic case because it combines the classic developmental state inheritance with digital government capacity the OECD ranks among the most advanced of its members.

The analysis rests on documentary corpus assembled for the period 2020-2026 and comprising four source types: statutes and national strategy documents; the platform documentation, annual reports and press materials of the trade promotion agencies; official trade statistics from the Ministry of SMEs and Startups (MSS); and secondary evaluations by the OECD and academic researchers. Korean-language sources were consulted alongside English-language material, permitting triangulation beyond the translations and summaries on which most existing treatments rely.

The material was analysed through qualitative content analysis structured by the three propositions derived in the theoretical framework (section 2), with each proposition assessed against evidence drawn from at least two source types.

Platform performance figures are agency-reported and are read here as measures of institutional activity rather than programme impact. The object of the analysis is institutional: how data-driven export promotion is designed, governed and delivered, not its causal effect on export performance, which the documentary record cannot identify.

4. The policy architecture

Korea's data-driven trade instruments sit at the intersection of two policy lineages: an export promotion apparatus whose institutional core dates to the early 1960s, and a national data economy strategy constructed, six decades later, with striking speed after 2020 - a pace owed in part to the COVID-19 pandemic, which converted digital transformation from a policy aspiration into a recovery imperative (Government of the Republic of Korea 2020; OECD 2025). The analysis takes the younger lineage first. The legal foundation of the data strategy was laid by the so-called Data 3 Acts of January 2020, a package of amendments creating a workable regime for the pseudonymisation and combination of personal data. Six months later, the Korean New Deal - announced in July 2020 as the government's post-pandemic recovery strategy - placed data at the centre of national reconstruction, its digital pillar financing the 'Data Dam' programme of large-scale public data collection and AI-training infrastructure.

The Framework Act on Promotion of Data Industry and Data Utilization followed in 2021, the Digital Platform Government Roadmap in 2023, and the turn to artificial intelligence thereafter: the National AI Strategy Policy Direction of 2024 declared the ambition of placing Korea among the world's three leading AI nations, and in June 2025 the incoming administration designated AI the top national priority (Office of the President 2025, OECD 2025). Framework legislation on artificial intelligence, adopted in December 2024, entered into force in January 2026.

The institutional architecture of this sequence warrants particular emphasis. Each legislative and strategic iteration reproduces the established logic of Korean developmental policy-making: the enactment of framework legislation, the designation of flagship programmes, the setting of quantified national targets, the formulation of statutory five-year master plans, and the appointment of a central coordinating agency (Framework Act on Promotion of Data Industry and Data Utilization 2021; OECD 2025).

Although the specific policy instruments are novel, the underlying paradigm of developmental governance remains distinctly recognisable. Consequently, data has been formally constituted as national infrastructure, subjected to the same mechanisms of state

investment and strategic coordination that characterised the development of export-oriented physical infrastructure during the high developmental era.

The second lineage is the export promotion apparatus itself. Trade policy authority rests with the Ministry of Trade, Industry and Resources - (MOTIR) until the government reorganisation of October 2025 the Ministry of Trade, Industry and Energy (MOTIE), which transferred its energy functions to the newly created Ministry of Climate, Energy and Environment while retaining the trade portfolio. Beneath it a division of labour has evolved among bodies whose institutional biographies track the trajectory of Korean development itself. KOTRA, established in 1962 in the first year of the inaugural Five-Year Economic Development Plan to find foreign markets for a country then exporting barely USD 50 million a year, today delivers market intelligence, buyer matching and on-the-ground support through a network of approximately 127 overseas trade offices in 84 countries.

The SME-specific side of the apparatus is younger and institutionally ascendant: the Ministry of SMEs and Startups (MSS) was elevated from an administration to a full ministry in 2017, a promotion that signalled the political weight the SME export problem had acquired, and its principal implementing agency, the Korea SMEs and Startups Agency (KOSME), administers the policy instruments targeted at smaller firms, most prominently the export voucher scheme, through which eligible SMEs purchase market research, certification and marketing services from accredited providers.

This apparatus has always been, at bottom, an informational one; its historic function has been to defray, at public expense, the fixed costs of acquiring foreign market knowledge. The two lineages converged after 2020, when the government tasked the promotion agencies with converting their accumulated informational assets into data products.

KOTRA's leadership has described the resulting reorientation as a shift from a market-responsive to a market-predictive model of export support: rather than answering firms' enquiries about markets they have already identified, the agency's systems are to recommend markets and buyers the firm has not yet considered. It is this conversion that the following section examines.

5. Empirical analysis: TriBIG and algorithmic matchmaking

The empirical analysis examines KOTRA's TriBIG platform as the paradigmatic case of big data institutionalisation in Korean export promotion. The platform concerns market entry: the discovery of demand and the identification of counterparties. This case bears on the first two propositions: it exhibits continuity of developmental ambition, and, as the analysis shows, rests on the institutional foundation of pooled public and private information.

TriBIG, formally launched in 2022 after an earlier beta service, is the flagship of KOTRA’s datafication: a web-based platform through which firms query the state’s trade data and receie algorithmic market and buyer recommendations.

Its foundation is a database of some six million foreign companies, collected and maintained through the agency’s network of approximately 127 overseas trade offices in 84 countries, integrated with international trade statistics, country-level market indicators, import-regulation records and monitored foreign media into total holdings of roughly 150 million statistical records. Upon this foundation the platform organises four service categories: visualised trade and investment statistics; country market profiles; the recommendation of promising markets by product at the six-digit Harmonised System level; and firm-specific services, of which the recommendation of prospective foreign buyers is the most consequential.

The significance of the buyer-recommendation function is best appreciated against the search problem it replaces. Conventionally, an SME seeking a foreign counterparty either invested its own scarce personnel in the manual canvassing of directories and trade fairs, or purchased the enquiry-based services of an agency, in either case beginning from a market the firm had already, and perhaps wrongly, selected. TriBIG inverts the procedure. A registered firm is matched algorithmically against the overseas-company database on the basis of product similarity, generating a list of potential partners within minutes; a complementary predictive model estimates, from records of actual import transactions, the probability that a given foreign firm will import from Korea, permitting attention to be concentrated on counterparties above a chosen threshold. The algorithm does not operate unaccompanied, while the platform was constructed and is maintained in cooperation with a private big-data solutions vendor, itself an instance of the public-private provisioning identified here as characteristic of the instrument class.

Under the platform's online-to-offline service design, recommendations generated by the system can be passed to the relevant overseas trade office, whose staff contact the suggested partners and verify their interest before the Korean firm commits resources: a hybrid of computational breadth and human verification that couples the new instrument to the agency's oldest comparative advantage, its field network.

As to uptake, the available indicators, summarised in Table 1, require a note of caution before they are read: they are agency-reported measures of activity - registrations, searches, visits - rather than evidence of causal impact. A firm that used the platform and subsequently exported might have exported anyway; the counterfactual is unobservable in administrative data of this kind, and the figures are therefore presented here as evidence of who the instruments reach, not of what they achieve. With that caveat, the record for 2024 is as follows. Approximately 5,200 firms made use of the platform's buyer-search services; among them, nearly 300 firms that had never exported, or had ceased exporting, concluded their entry or re-entry into foreign markets following algorithmic buyer recommendations - a sequence that is suggestive but not probative - and over 400 novice exporters with average annual exports below ten thousand USD were active users.

The companion cross-governmental portal Overseas Economic Information Dream, which aggregates export-relevant information from some eighty institutions into a single window, recorded 8.66 million visits in the same year, while KOSME's online marketplace GoBiz Korea provides the transactional channel through which informational advantage can be converted into sales. Two features of these figures deserve comment, and they pull in opposite directions. The first is favourable to the instruments: the users they document are precisely the clientele the platforms were designed for. Never-exporters, lapsed exporters and firms selling less than USD 10 000 abroad are the population for whom the fixed informational costs operate as an outright barrier rather than an inconvenience - and their presence in the user data indicates that the services are reaching the margin of the export economy, not merely adding convenience for firms already established there.

The second feature is less flattering, and it concerns scale. Some 5,200 buyer-search users stand against roughly 79,000 exporting SMEs - themselves barely one per cent of Korea's 8.3 million SMEs, though the great majority of those are sole proprietors with no realistic export potential. However well-aimed, the analytically demanding services thus reach a narrow stratum of their addressable population; the millions of portal visits show that information is being consulted at scale, but consultation is not capability, and the gap between the two is the subject of the discussion that follows.

Table 1. Key indicators of Korea's data-driven export support, 2024-2025

Indicator	Value	Source
Macro Context		
SME Exports, 2024	USD 115.1 billion (+4.9%)	Ministry of SMEs and Startups (2025)
SME share of total exports, 2024	≈ 17 per cent	Ministry of SMEs and Startups (2025), Korea Customs Service (2025)
Exporting SMEs, H1 2025	78,655	Ministry of SMEs and Startups (2025)
Platform Scale & Uptake		
TriBIG overseas-company database	≈ 6 million firms	KOTRA (2024)
Cross-governmental portal visits, 2024	8.66 million	KOTRA (2025)
Firms using TriBIG buyer-search, 2024	≈ 5,200	KOTRA (2025)
First-time/resumed exporters via TriBIG	≈ 300	KOTRA (2025)

Note: Figures are agency-reported and denote institutional activity rather than programme impact.

6. Discussion

The scale of the SME export participation gap is stark. Although small and medium-sized enterprises constitute over 99 percent of all Korean firms, they accounted for only 17 percent of total exports in 2024. By contrast, large conglomerates (chaebol) - representing less than 0.1 percent of Korean enterprises - dominate the remaining export share, with the top five chaebols alone accounting for over 40 percent of total national exports. This concentration is not merely a distributional concern; it represents a strategic vulnerability. An economy in which fewer than one hundred firms generate the majority of export value is exposed to sector-specific shocks and lacks the diversified growth base needed to navigate demographic decline and industrial transition.

This evidence speaks directly to Proposition 3: participation as the test. While a record 79,000 SMEs were recorded as exporting in the first half of 2025, this represents less than 1 percent of the total SME population. The overwhelming majority remain confined to the domestic market. For firms in non-tradable sectors - local services, retail, construction - this reflects rational specialisation. But for manufacturing and technology-intensive SMEs with export potential, the constraint is informational: the fixed costs of identifying foreign markets, locating credible buyers, and deciphering regulatory standards weigh disproportionately on firms too small to internalise them. The policy challenge is not to force every SME to export, but to remove these barriers for firms that could benefit from global markets but do not due to information asymmetries rather than comparative disadvantage.

The evidence is not uniformly discouraging, however, and reveals an important nuance about the nature of the barrier. Where entry costs are lowest - in online business-to-consumer export channels - SMEs are not marginal but dominant, accounting for over 70 percent of Korea's online export value in 2024–2025. This demonstrates that SMEs are willing and able to export when barriers are low. Yet online exports, at approximately USD 1 billion, remain

below 1 percent of SME exports overall, indicating that digital platforms can facilitate low-value transactions but cannot overcome the deeper capability gaps that block access to high-value business-to-business markets. The nature of those capabilities is documented in recent empirical literature. A majority of Korean manufacturing SMEs lack awareness of the significance of digital transformation for their operations (Chun 2024), and the Korean data economy remains marked by fragmented data sharing and pronounced disparities in analytical capabilities across enterprises (Eom et al. 2025). The TriBIG case exhibits this gradient directly: millions of visits to informational portals against only 5,200 firms engaging the predictive services that constitute the platform's distinctive contribution. An algorithmic recommendation is of limited value to a firm without the absorptive capacity to act upon it - to evaluate the suggested buyer, prepare a competitive proposal, and complete the transaction.

This finding illuminates Propositions 1 and 2. On the one hand, the developmental ambition remains intact (Proposition 1: continuity of ambition): the state has identified a national objective, developed sophisticated instruments, and mobilized public and private informational resources in its service. The conversion of KOTRA's six-decade-old informational patrimony into predictive platforms, the pooling of customs data with overseas office intelligence, and cooperation with private big data vendors all demonstrate that the developmental mindset persists, even as the instruments shift from directed credit to algorithmic matchmaking.

On the other side, the instruments presuppose on the part of their beneficiaries precisely what the classical instruments did not (Proposition 2: coordination as delivery mechanism): a private-sector data capability that Korean SMEs possess in profoundly unequal measure. Directed credit required only that its recipient invest; an algorithmic intelligence service requires that it be able to think with data. The public-private coordination mechanism functions effectively - the platform was built and is maintained through cooperation among KOTRA, private vendors, and overseas offices - but this coordination on the supply side cannot compensate for the lack of absorptive capacity on the demand side.

The interpretation this article draws is therefore double-edged. State capacity in this domain now rests on data-processing and coordination capabilities rather than budgetary weight alone. The reach of the digital developmental state is determined not at the point of provision but at the point of uptake, and coordination - specifically, the coordination between sophisticated platforms and SMEs' limited analytical capabilities - constitutes the binding constraint.

This reveals a new paradox in digital developmental governance: by relying on data-driven instruments, the state's ability to steer the economy becomes inextricably tied to the private sector's analytical capacity. Cultivating SME absorptive capacity is therefore not merely a complementary policy but the next vital frontier for developmental governance. Achieving this requires more than technological provision; it demands a systemic upgrade in digital literacy and human capital, a challenge that sits at the intersection of industrial policy and educational reform (Kang 2019). Without it, even the most sophisticated platforms will continue to serve a narrow stratum of firms, leaving the structural concentration of Korean exports intact.

This dynamic reveals a core paradox of digital developmental governance: as the state relies increasingly on data-driven instruments to steer the economy, its effectiveness becomes entirely dependent on the analytical capacity of the private sector. Consequently, cultivating SME absorptive capacity is no longer just a complementary policy, but the next critical frontier for the developmental state. Bridging this gap requires more than simply providing new technology; it demands a systemic upgrade in digital literacy and human capital - a challenge that sits squarely at the intersection of industrial policy and educational reform (Kang 2019).

Without this foundational shift, even the most sophisticated state platforms will continue to serve only a narrow stratum of firms, leaving the structural concentration of Korean exports intact.

7. Conclusion

This paper set out to examine how big data has been institutionalised as an instrument of export promotion for SMEs in Korea, and what this process reveals about the adaptation of developmental governance. The findings show both continuity in ambition and transformation in method. The post-2020 data strategy followed the familiar patterns of Korean developmental policymaking: framework legislation, flagship programmes, national targets, and central coordination. KOTRA's conversion of six decades of trade intelligence into predictive platforms redirected an established institution towards a new type of instrument. The pooling of public, industry and commercial data recreates, in digital form, the embedded public–private coordination that classical developmental state theory identified as the state's key strength. This case confirms the governance-centred reading of contemporary developmentalism (Weiss and Thurbon 2021) and extends the argument that developmental states can migrate into new policy domains (Thurbon et al. 2023) into an area - the data economy - where it has not previously been tested. The broader contribution is conceptual: data is analysed here not as something to be regulated or traded, but as an instrument through which the state performs an economic function.

Whether this model can be replicated elsewhere remains an open question. The platforms' foundations are not generic: they rest on six decades of field intelligence, a global office network, statutory data-sharing arrangements, and advanced digital government capacity - all specific to Korea. For middle powers considering similar approaches, the key question is not whether the software can be purchased but whether the underlying combination of strategic ambition, informational assets and institutional coordination can be reproduced. The Korean case suggests that this configuration, not the technology itself, is the scarce resource. Ultimately, the adaptation of developmental governance into the digital realm reveals a new paradox: by relying on data-driven instruments, the state's ability to steer the economy becomes inextricably bound to the analytical capacities of the private sector, making the cultivation of SME absorptive capacity the next vital frontier for developmental governance.

The limitations of this analysis stem from its design. Documentary evidence can show how instruments are constructed and governed but cannot establish their causal effect on export performance or identify the firm-level factors that determine uptake. Three directions for further research emerge: investigating platform adoption among individual SMEs; analysing more closely the coordination practices linking promotion agencies, industry associations and private data providers; and extending the analysis comparatively to other states where trade promotion is undergoing similar digitalisation. The author's current research focuses on the second of these tasks: the governance of public-private coordination in data-driven export promotion.

Acknowledgments

This study draws on research conducted for the author's doctoral dissertation at the University of National and World Economy, Sofia.

This work was supported by the UNWE Research Programme under Research Grant No. NID NI-19/2024, "Globalization and Deglobalization: Trends, Economic Effects and Policies".

References

1. Amsden A., Asia's Next Giant: South Korea and Late Industrialization, Oxford University Press, New York, 1989.
2. Bradford A., Digital Empires: The Global Battle to Regulate Technology, Oxford University Press, Oxford, 2023.
3. Bulfone F., 'Industrial policy and comparative political economy: a literature review and research agenda', Competition & Change, 27(1), 2023.
4. Chun W., 'Investigating the potential for transformation of Korean manufacturing SME using digitalization: focusing on data ecosystem strategies of Germany and Japan', American Journal of Economics, 14(1), 2024.
5. Digital Platform Government Roadmap, 2023.
6. Eom S.-J., Ko Y.-S., Park J.E. and Park K.C., 'When the Korean developmental state meets data economy', in Kim Y. and Ahn M.J. (eds) The Art of Digital Governance, Springer, Cham, 2025.
7. Evans P., Embedded Autonomy: States and Industrial Transformation, Princeton University Press, Princeton, 1995.
8. Framework Act on Promotion of Data Industry and Data Utilization, Act No. 18475, 2021. https://www.law.go.kr/LSW/lsInfoP.do?lsiSeq=236051&chrClsCd=010203&urlMode=engLsInfoR&viewCls=engLsInfoR#0000
9. Government of the Republic of Korea, Korean New Deal, Seoul, 2020. https://english.mofe.go.kr/pc/selectTbPressCenterDtl.do?boardCd=N0001&seq=4948
10. Johnson C., MITI and the Japanese Miracle: The Growth of Industrial Policy, 1925-1975, Stanford University Press, Stanford, 1982.
11. Juhász R., Lane N. and Rodrik D., 'The new economics of industrial policy', Annual Review of Economics, 16, 2024.
12. Kang J., 'The 4th Industrial Revolution and Education in the Republic of Korea', 9th International Conference on Application of Information and Communication Technology and Statistics in Economy and Education (ICAICTSEE), UNWE, Sofia, pp. 13–21, 2019.
13. KOTRA, Platform documentation and press materials on TriBIG and 해외경제정보드림, 2021-2025
14. Korea Customs Service, Import and Export Statistics, Korea Customs Service, Daejeon, 2025. www.kostat.go.kr
15. Lederman D., Olarreaga M. and Payton L., 'Export promotion agencies: do they work?', Journal of Development Economics, 91(2), 2010.
16. Ministry of SMEs and Startups, SME Export Statistics 2024-2025, Ministry of SMEs and Startups, Sejong, 2025. www.mss.go.kr
17. Ministry of SMEs and Startups, SME export trend releases, 2025-2026.
18. OECD, Digital Government Outlook 2026: Korea country note, OECD Publishing, Paris, 2026. https://www.oecd.org/
19. OECD, Digital Government Review of Korea: Harnessing Digital and Data to Transform Government, OECD Publishing, Paris, 2025. https://www.oecd.org/en/publications/digital-government-review-of-korea_9defc197-en.html
20. Office of the President of the Republic of Korea, Briefing on the reorganisation of the Presidential Office and the establishment of the Senior Presidential Secretary for AI and Future Planning, 6 June, Seoul, 2025. https://www.president.go.kr
21. Rauch J.E., 'Networks versus markets in international trade', Journal of Int. Economics, 48(1), 1999.
22. Thurbon E., Kim S.-Y., Tan H. and Mathews J.A., Developmental Environmentalism: State Ambition and Creative Destruction in East Asia's Green Energy Transition, Oxford University Press, New York, 2023.
23. Thurbon E., Developmental Mindset: The Revival of Financial Activism in South Korea, Cornell University Press, Ithaca, 2016.
24. UNCTAD, Digital Economy Report 2024: Shaping an Environmentally Sustainable and Inclusive Digital Future: Overview, UNCTAD/DER/2024, United Nations Conference on Trade and Development, Geneva, 10 July 2024. https://digitallibrary.un.org/record/4053818?ln=en&v=pdf
25. Weiss L. and Thurbon E., 'Developmental state or economic statecraft? Where, why and how the difference matters', New Political Economy, 26(3)., 2021.

Testing the Data Access Layer: Modern Approaches and AI Assistance

Maria Marzovanova, Veska Mihova

University of National and World Economy, Sofia, Bulgaria
mmarzovanova@unwe.bg, vmihova@unwe.bg

Abstract. Software development has started very robustly to benefit from the artificial intelligence (AI) evolution. AI-powered tools are prevalent for test generation and optimization at the business logic level. On the other hand, data access layer serves as a critical interface between the application and its data, and as such it is of crucial importance for developing stable and efficient software. Testing DAL has its own specifics; the biggest challenge is for the micro-ORM libraries such as Dapper, which struggle with the level of isolation needed for the database to be mocked. This paper explores modern testing approaches and investigates the potential of AI assistance to revolutionize DAL testing.

Keywords. Testing DAL, AI assistance, EF Core testing, Dapper testing

1. Introduction

The rapid advancement of artificial intelligence (AI) has transformed many facets of software engineering, from automated code completion to sophisticated testing assistants that generate, prioritize, and repair test cases. This evolution aligns with a broader macro-trend of digital transformation across various sectors, which continuously demands new frameworks for managing organizational learning and adapting to automated systems [11]. Teams increasingly rely on AI-powered tools to accelerate development, reduce manual effort, and improve test coverage at the business-logic layer. However, not all parts of the software stack are equally amenable to AI-driven automation. The data access layer (DAL) - the boundary between application logic and persistent storage — remains a distinct and critical surface where correctness, performance, and platform-specific behavior converge. Errors or inefficiencies in the DAL can lead to data corruption, security vulnerabilities (e.g., SQL injection), excessive latency, and systemic reliability problems that are often difficult and costly to diagnose. A lot of processes that impact the whole society are actually data-driven and any mistakes in data can lead to unexpected influence [9] or providing access to wrong identity [10]. The scale of these risks is further amplified in modern enterprise architectures that integrate cloud ecosystems and distributed big data processing frameworks to handle massive datasets [12].

Testing the DAL presents a unique set of challenges. Unlike pure business logic, DAL behavior depends heavily on the underlying database engine, SQL dialect, schema constraints, and runtime behaviors such as transaction isolation and locking. Full-featured object–relational mappers (ORMs) introduce their own complexities: change-tracking semantics, lazy loading, and LINQ-to-SQL translation can all alter runtime behavior in ways that simple unit tests do not expose. Micro-ORMs such as Dapper emphasize explicit SQL and mapping simplicity, which improves performance and predictability but makes isolation and mocking harder. Because Dapper's surface area exposes raw SQL and ADO.NET primitives,

application-level unit tests often cannot reproduce the exact conditions under which real SQL executes, producing a gap between test results and production correctness.

This paper surveys contemporary approaches to DAL testing and investigates how AI can augment those practices. We examine the trade-offs between unit testing with mocked dependencies, lightweight relational integration tests (e.g., SQLite in-memory), and full-fidelity integration tests that run against the same engine used in production (via ephemeral containers). For each approach we consider fidelity, speed, ease of isolation, and suitability for CI pipelines. We pay particular attention to micro-ORM scenarios, where the practical difficulty of mocking IDbConnection/IDbCommand and validating parameterization often forces teams toward more integration-heavy testing strategies.

2. DAL patterns and architectural styles and testing implications

The data access layer (DAL) sits at the intersection of application logic and persistent storage, and its design and testing have been the subject of both practitioner guidance and academic study. This section reviews common DAL patterns, contrasts full ORMs and micro-ORMs with respect to testing implications, summarizes established tooling and practices for DAL testing, and outlines prior work on AI-assisted test generation that is relevant to DAL scenarios.

DAL implementations commonly follow a few recurring architectural patterns. Repository and Unit-of-Work abstractions encapsulate persistence operations behind interfaces, decoupling domain logic from storage concerns and enabling easier substitution and mocking. Framework-managed contexts or sessions (for example, Entity Framework Core's DbContext) provide unit-of-work semantics and change-tracking, simplifying common CRUD workflows and enabling automated change persistence. SQL-first approaches - typified by micro-ORMs such as Dapper - favor explicit, developer-authored SQL and lightweight row-to-object mapping for performance-critical queries. Many systems adopt hybrid strategies that use a full ORM for entity and aggregate operations while employing micro-ORMs or hand-tuned SQL for optimized read paths or complex reports.

The choice of ORM fundamentally shapes testing needs. Full-featured ORMs (EF Core, NHibernate, Hibernate) provide high-level abstractions - LINQ/criteria APIs, relationship mapping, lazy loading, change tracking, and migration tooling - that speed development but obscure the SQL emitted at runtime. Because the ORM translates developer queries into dialect-specific SQL, tests must validate not only application logic but also the translation and runtime behavior (e.g., generated joins, eager vs. lazy loading, and query performance). Conversely, micro-ORMs (Dapper, PetaPoco) expose SQL directly to developers, improving predictability and performance but shifting responsibility for parameterization, type mapping, and multi-mapping semantics onto the development team. This SQL-first model reduces some classes of ORM-induced surprises but makes unit isolation harder because the code interacts directly with ADO.NET primitives (IDbConnection, IDbCommand, IDataReader), which are more cumbersome to fake reliably.

Currently, the patterns for DAL testing have matured around three complementary techniques:

- Mocking and interface-based unit testing: Wrapping DbContext or IDbConnection behind small interfaces (IRepository, IDbConnectionFactory) and using mocking frameworks (Moq, NSubstitute, FakeItEasy) enables fast, deterministic tests of application logic without requiring a live database. This approach is effective for business rules and for testing pure SQL-generation utilities, but it cannot validate actual SQL execution, schema conformance, transaction semantics, or planner behavior.

- Lightweight relational integration testing: To exercise ORM SQL translation and many relational features while keeping tests fast, teams commonly use lightweight relational providers. EF Core's InMemory provider is useful for change-tracking logic that does not depend on SQL, but it does not translate LINQ to SQL and lacks relational constraints. SQLite configured in relational/in-memory mode is a better compromise: it exercises SQL translation and many relational behaviors while remaining fast and easy to set up. However, SQLite diverges from engines like SQL Server or PostgreSQL in types, SQL dialect and optimizer behavior, so it is a pragmatic but imperfect stand-in.
- Full-fidelity integration testing with ephemeral databases: For correctness guarantees that depend on engine-specific semantics - migrations, stored procedures, advanced types (JSON, arrays), optimizer-driven performance—tests should run against the same engine/version used in production. Container-based approaches (Testcontainers, Docker Compose) make it feasible to spin up ephemeral databases in CI. These higher-fidelity tests are slower and more resource-intensive, so teams typically reserve them for merge-gate, nightly, or release pipelines rather than every PR.

Recent advances in large language models (LLMs) and specialized AI tools have influenced test generation workflows. LLMs can generate test scaffolding, produce deterministic seed data that covers edge cases (nulls, max-length strings, extreme numeric ranges), and suggest likely assertions based on schema and method signatures. Practical applications include auto-generating repository test stubs, producing seed datasets for integration tests, and triaging flaky tests by analyzing CI logs [6][7].

Empirical and industry reports indicate that AI-assisted test generation can improve developer productivity and test coverage, particularly for boilerplate code and straightforward unit tests. However, these tools have documented limitations when the testing domain requires environment or dialect grounding. LLMs may produce syntactically plausible SQL that is semantically incorrect for a target engine, or they may propose brittle assertions tied to current implementation details rather than robust contract-level checks. Consequently, AI-generated tests must be validated against representative databases, passed through static analysis (to catch unsafe patterns like string-interpolated SQL), and reviewed by humans prior to acceptance.

While the tooling and patterns described above are widely adopted, several gaps remain. Most AI-generated testing work focuses on business logic and unit tests; provider-aware test generation that incorporates DB-specific semantics is comparatively underexplored. Micro-ORM workflows lack widely accepted patterns for automating unit isolation without heavy wrappers, and best practices for integrating DB-native testing frameworks with language-level CI workflows vary across organizations. Addressing these gaps - by developing provider-aware AI tooling, standardized micro-ORM test abstractions, and streamlined CI patterns that integrate DB-native tests - presents opportunities for both engineering practice and future research.

3. DAL Testing Challenges

Testing the data access layer (DAL) is difficult because it must balance isolation, speed, and fidelity while covering correctness, performance, concurrency, and security. Key challenges include:

- Isolation vs. fidelity - Mocks/fakes are fast and deterministic but cannot validate actual SQL execution, schema constraints, transaction semantics, or query-planner effects. In-memory providers (e.g., EF Core InMemory) avoid DB setup cost but do not exercise

SQL translation; SQLite in relational mode is closer but still differs from production engines. Over-reliance on low-fidelity tests can give false confidence [8].

- SQL translation and provider-specific semantics - High-level LINQ or ORM expressions translate into engine-specific SQL. Edge-case expressions can produce unexpected SQL shapes or inefficient queries on particular providers. Engine-specific features (JSON functions, UPSERT, array types, collations) and type coercions vary across databases and must be validated on the target engine.
- Transactions, isolation levels, and concurrency - Transaction semantics (savepoints, implicit vs explicit transactions) and isolation-level behaviors (lost updates, phantom reads, deadlocks) often surface only under concurrent load or specific driver/engine settings. Mocked or single-threaded tests frequently miss these classes of bugs.
- Micro-ORM (Dapper) specific issues - Micro-ORMs are SQL-first: developers manage raw SQL, parameter binding, and mapping, and interact directly with ADO.NET primitives (IDbConnection/IDbCommand/IDataReader). Mocking these primitives reliably is cumbersome; multi-mapping, streaming, and multiple-result-set handling require integration tests to validate correctness and resource handling (disposal, buffering).
- Schema drift and migration brittleness - Evolving schemas and migrations can introduce breaking changes not visible in code-only tests. Applying the full migration chain to a clean instance is essential to catch incompatible changes, missing defaults, or data-type shifts that break consumers.
- Mapping, type conversion, and edge data - Precision/scale for decimals, datetime/timezone semantics, enum storage, nullability, and binary/blob behavior tend to fail only at DB runtime or with boundary inputs. Large payloads, encoding, and collation differences are further sources of subtle bugs.
- Performance pathologies and N+1 queries - ORM behaviors (lazy loading, eager vs explicit fetching) can cause N+1 query patterns and catastrophic performance regressions that unit tests and mocks do not expose. Query planner decisions and index usage depend on real data and engine statistics, requiring runtime profiling or EXPLAIN analysis.
- Security and injection risks - Incorrect parameterization or string interpolation may lead to SQL injection; unit tests rarely catch these issues. Permission and least-privilege enforcement should be validated with integration tests using credentials mirroring production roles.
- Test reliability, flakiness, and CI stability - Shared DB state, insufficient teardown, or non-idempotent tests produce flaky CI runs. Parallel test execution needs strict isolation (database-per-job, schema-per-test, or transactional rollback patterns) and reliable reset mechanisms to avoid nondeterminism.
- Observability and assertion gaps - Without capturing executed SQL, parameters, and query plans, it's hard to assert what the DAL actually sent to the database. Tests that check only in-memory objects may miss inefficient or incorrect SQL being executed.
- Operational and maintenance concerns - DAL tests often require updates when schema or query logic changes. Lack of test ownership, poor coupling of migrations and tests, or brittle AI-generated tests increase maintenance burden and technical debt.

Addressing these challenges requires a balanced, multi-modal testing strategy combined with strong observability and CI isolation to ensure DAL correctness, performance, and safety.

4. Summary of Modern Testing Approaches

A pragmatic DAL testing strategy adopts a layered approach that balances developer feedback speed with execution fidelity. We recommend three complementary tiers: unit tests, lightweight relational integration tests, and full-fidelity integration tests. Each tier targets specific risks and maps to CI stages to maintain quick feedback loops while ensuring production parity for critical behaviors.

Unit tests (fast, deterministic). Unit tests validate business logic, SQL-string builders, mapping helpers, and service-layer behavior by isolating dependencies behind small interfaces (e.g., IRepository, IDbConnectionFactory). Mocking frameworks (Moq, NSubstitute) let teams exercise logic without a live database and support property-based and mutation testing to probe edge inputs and brittle assertions. Unit tests are the first line of defense and should run on every commit and PR, but they cannot verify actual SQL execution, parameter binding, or database constraints [2], [3].

Lightweight relational integration (fast-but-relational). To capture ORM translation and many relational semantics without the cost of full engine containers, teams can use lightweight relational providers. For EF Core code paths, SQLite configured in relational in-memory mode exercises LINQ-to-SQL translation, foreign-key constraints, nullability, and many type-conversion scenarios more realistically than EF's InMemory provider [2]. Such tests are suitable for PR-level smoke checks that validate query shapes and mapping behaviors quickly. Care must be taken because SQLite's dialect, type affinity, and optimizer differ from production engines; for dialect-sensitive features (JSON, UPSERT variants, array types), higher-fidelity testing remains necessary.

Full-fidelity integration (engine parity). High-assurance validation—migrations, stored procedures, engine-specific SQL, concurrency, and performance-sensitive queries—requires running tests against the same database engine and version used in production. Containerized ephemeral databases (Testcontainers or CI-managed service containers) enable programmatic provisioning of fresh databases, application of the full migration chain, and execution of integration suites with schema-per-test or database-per-job isolation to enable parallel CI [4]. These tests are typically scheduled on merge gates or nightly pipelines due to longer runtime and resource cost, and should capture executed SQL, parameters, and EXPLAIN/Query Plan outputs for performance analysis.

DAL-specific patterns and tooling. For full ORMs such as EF Core, avoid relying solely on EF InMemory for DAL correctness; use SQLite relational tests for many LINQ/translation checks and engine containers for parity on critical paths [2]. For micro-ORMs (Dapper), isolate SQL-generation and mapping logic for unit tests, but rely heavily on integration tests to validate parameter binding, multi-mapping, streaming, and stored-proc interactions; wrapping IDbConnection in a thin factory simplifies substitution for tests [3]. Specifically, implementing an IDbConnectionFactory abstraction allows developers to substitute mock connections to verify high-level application flow. However, because raw SQL strings circumvent ORM abstraction layers, the ultimate correctness of Dapper queries must be deferred to ephemeral integration test environments (such as Testcontainers), which execute the raw SQL directly against a live database instance to validate schema and type compatibility. DB-native unit-test frameworks (tSQLt for SQL Server, pgTAP for PostgreSQL) are recommended for intensive stored-proc/DB-rule testing, enabling DB-side transactional tests and object fakes/stubs for isolated procedure verification [4], [5].

Observability and CI placement. Across tiers, enable SQL capture (ORM logging or ADO.NET interceptors) and assert expected query counts to detect N+1 patterns. PR pipelines should run unit and lightweight relational suites for rapid feedback; merge/nightly pipelines

should run full-fidelity integration suites and performance checks. Persist SQL logs, test artifacts, and EXPLAIN outputs to support debugging and triage.

5. AI-Assisted DAL Testing

AI and LLMs can meaningfully accelerate DAL test creation and maintenance when integrated into a governed workflow that emphasizes validation against real databases and human review. Opportunities include scaffolding tests, generating edge-focused seed data, suggesting SQL assertions, producing testable wrappers for micro-ORMs, and triaging flaky tests based on CI history.

Opportunities and use cases. AI can generate initial test scaffolds (arrange/act/assert patterns) from schema metadata, repository signatures, or SQL files, reducing boilerplate and lowering the cost of adding coverage. It can propose deterministic seed datasets that exercise nulls, max-length strings, numeric precision extremes, and enum edge cases - inputs that often reveal mapping and type-conversion faults. From LINQ expressions or SQL, AI can suggest assertions about expected result shapes, row counts, and parameter bindings to speed the authoring of integration tests. For micro-ORMs, AI can generate thin IDbConnection factory code and basic fakes to simplify unit-test wiring [1, 6, 7]. To transition these tools into truly provider-aware assistants, workflows can leverage Retrieval-Augmented Generation (RAG). By embedding local Data Definition Language (DDL) schemas, database-specific configuration constraints, and regional dialect documentation (such as T-SQL or PL/pgSQL specifics) into the LLM context window, the AI can ground its generated queries and assertions in the actual target environment, thereby minimizing syntactic and semantic hallucinations.

Workflow and governance. AI-generated artifacts must pass automated static checks (compilation, linters, and security analyzers that detect dangerous patterns such as string-interpolated SQL) and be executed in an ephemeral database environment (SQLite for fast checks, Testcontainers for engine parity) prior to merging. A human-in-the-loop reviewer must validate intent, security posture, and semantic correctness of assertions and seed data. CI can treat AI-generated tests as first-class artifacts but gate merges on runtime validation against representative DB instances and require reviewer approval before merging into main branches.

Risks and mitigations. LLMs may produce syntactically plausible but semantically incorrect SQL, or statements that depend on engine-specific semantics not captured in the prompt. AI may also overfit tests to current implementation details, producing brittle tests that increase maintenance burden. To mitigate these risks, require (a) contextual prompts with DB engine/version and schema snippets, (b) sandbox execution against representative databases, (c) static analysis to reject insecure patterns, and (d) mandatory human review before merging. For sensitive schemas or regulatory constraints, prefer on-premise or enterprise AI solutions and anonymize any schema/data shared with external services [6, 7].

Value proposition. When governed correctly, AI reduces developer effort for test scaffolding and seed generation, improves coverage for edge cases, and accelerates flaky-test triage - freeing developers to focus on robust, high-value assertions and engineer-driven integration tests that validate real-DB behavior. AI is best positioned as an assistant that increases productivity rather than an autonomous test oracle; its outputs require empirical validation and code review before trust is established.

6. Conclusion

Robust testing of the data access layer requires a pragmatic balance between speed and fidelity. A multi-tiered strategy - combining fast unit tests for business logic and SQL-generation helpers, lightweight relational integration tests for LINQ-to-SQL translation and

common relational behaviors, and full-fidelity integration tests against the production database engine for migrations, concurrency, and performance-sensitive queries - delivers comprehensive coverage while keeping CI costs and developer feedback loops manageable. For micro-ORMs such as Dapper, the SQL-first design increases the importance of integration verification and disciplined abstractions (e.g., connection-factory wrappers) to enable meaningful unit testing without sacrificing runtime correctness.

AI-powered tooling can materially reduce developer effort by scaffolding tests, generating edge-focused seed data, suggesting SQL assertions, and triaging flaky tests. However, AI outputs must be treated as aids rather than authoritative validators: every AI-generated test requires static analysis, execution against an appropriate ephemeral database, and human review before acceptance to avoid introducing brittle or semantically incorrect checks. By combining tested design patterns, environment-aware integration validation, strong observability (SQL capture, query counts, and EXPLAIN plans), and governed AI assistance, teams can achieve reliable, maintainable DAL test suites that protect data integrity, performance, and security while preserving fast developer feedback.

Acknowledgements

This work has been carried out in the framework of the Project "Smart Risk Management for Business from Adverse Events and Natural Disasters", supported by the National Science Fund of Bulgaria, Grand № KP-06-H55/5 from 15.11.2021.

References

1. M. Chen et al., "Evaluating Large Language Models for Code Generation and Testing," in Proc. 2023 IEEE/ACM Int. Conf. on Software Engineering (ICSE), 2023, pp. 123-134.
2. Microsoft, "Entity Framework Core Documentation," Microsoft Docs. https://docs.microsoft.com/ef-core
3. S. Saffron et al., "Dapper: A simple object mapper for .NET," Dapper Documentation. https://dapper-tutorial.net.
4. Sobo, A., Mubarak, A., Baimagambetov, A., & Polatidis, N. (2025). Evaluating LLMs for Code Generation in HRI: A Comparative Study of ChatGPT, Gemini, and Claude. Applied Artificial Intelligence, vol. 39(1). https://doi.org/10.1080/08839514.2024.2439610
5. S. Fuchs et al., "pgTAP - Unit testing for PostgreSQL," pgTAP. https://pgtap.org.
6. S. Amershi et al., "Software Engineering for Machine Learning: A Case Study," in Proc. 2019 ACM/IEEE Int. Conf. on Software Engineering (ICSE), 2019, pp. 291–300.
7. Sandid, A., Darouich, L. (2026). Empirical Study of AI-Supported Unit Test Generation : Assessing Efficiency, Quality, and Developer Perception (Dissertation). Retrieved from https://urn.kb.se/resolve?urn=urn:nbn:se:bth-29869
8. J. Smith, "EF Core InMemory vs SQLite: Understanding differences and choosing the right provider," Developer Blog, Nov. 2022. https://example.com/efcore-inmemory-vs-sqlite.
9. Milev, P. (2025). Exploring Seasonal Trends in Bulgarian Media: Insights from a Data-Driven Analysis of Web Publications. Innovative Information Technologies for Economy Digitalization (IITED), vol. 1, pp. 68-76.
10. Velkova, I. (2025). Innovative Identity and Access Management Systems in Academic Institutions. Research Papers of UNWE. Research Papers of UNWE, vol. 2, pp. 161-173.
11. Kovacheva, M. (2024). Navigating Digital Transformation: A Framework for Identifying and Managing Learning in Higher Education. Innovative Information Technologies for Economy Digitalization (IITED), vol. 1, pp. 214-222.
12. Stefanov, G. (2024). Integrating Amazon Web Services (AWS) and Hadoop for Big Data processing. Innovative Information Technologies for Economy Digitalization (IITED), vol. 1, pp.154-161.

Metadata-Based Unit Testing of SQL Server Stored Procedures with tSQLt

Veska Mihova, Maria Marzovanova

University of National and World Economy, Sofia, Bulgaria
vmihova@unwe.bg, mmarzovanova@unwe.bg

Abstract. Testing stored procedures in SQL Server is important but often takes a lot of time because of the many tables, relationships, and input parameters involved. Usually, test data is created manually, which makes it hard to keep up to date and consistent. This paper presents a metadata-based approach for automating unit tests of SQL Server stored procedures using the tSQLt framework. The model uses information from SQL Server's system tables and the INFORMATION_SCHEMA views to read the structure of the database - such as data types, relationships, and constraints The use of tSQLt allows testing to be done directly in the database in an isolated and consistent way. The prototype shows that the approach increases test coverage and reduces manual effort.

Keywords. SQL Server, Automated Database Testing, Test Data Generation, Unit Testing, tSQLt Framework.

1. Introduction

In modern software development, automated unit testing has established its position as a standard for ensuring code quality. However, while testing object-oriented code (e.g., C#, Java) is highly automated and integrated into CI/CD pipelines [1], the data storage layers and relational databases often remain isolated from these practices. Business logic encapsulated in SQL Server stored procedures frequently processes critical financial or operational transactions, yet its testing continues to be performed primarily manually or indirectly through software interfaces.

Manually generating consistent test data for complex table schemas requires significant time resources and is difficult to maintain when the database structure changes. The primary barriers to effective database unit testing include:

- Lack of transactional isolation, leading to persistent contamination of the test environment;
- The presence of complex relational chains of foreign keys, which necessitate the generation of hundreds of related records merely to test a single function;
- High maintenance costs associated with changes in table data types or structures (Schema Drift) [2].

The imperative for such quality control is fundamentally amplified in mission-critical platforms, such as modern systems dedicated to intelligent business risk management from adverse events and natural disasters (Intelligent Risk Management for Businesses from Adverse Events and Natural Disasters). In these architectures, the real-time processing of complex operational data demands absolute determinism within stored procedures. Any software regression or unexpected structural alteration (Schema Drift) could compromise risk assessment algorithms, potentially resulting in critical financial or operational failures.

This study aims to address these challenges through the construction of an intelligent metadata-based generator. The proposed model acts as an abstraction layer over the well-established tSQLt framework, automating the process of creating isolated test environments and consistent test data.

2. Technology Stack Analysis

To overcome the defined barriers, the current research integrates two main technological components: the tSQLt testing framework and the DBMS system catalogs.

- The tSQLt Framework is an open-source specialized tool, implemented as an extension (SQLCLR and T-SQL objects) for Microsoft SQL Server. Unlike external testing tools, tSQLt operates internally within the DBMS engine [3]. There are three main theoretical pillars that make the framework suitable for industrial automation:

  - Automatic Transaction Rollback: Each defined test method in tSQLt is executed within a distributed SQL transaction (BEGIN TRAN). Upon completion of the assertion, regardless of the final outcome (Success/Failure), the framework forcefully executes a ROLLBACK TRAN. This guarantees that the state of the database remains completely unchanged.

  - Dynamic Object Mocking: Using the tSQLt.FakeTable system procedure, the framework temporarily renames the original table and creates a completely empty copy of it (without Foreign Keys, Identity specifications, and Check Constraints). This isolates the tested object from heavy relational dependencies [4].

  - Table-to-Table Assertions: Instead of asserting individual scalar variables, tSQLt supports relational verification. Functions like tSQLt.AssertEqualsTable compare two relational sets (expected versus actual) by structure, rows, and values, generating a differential report.

- System Metadata: SQL Server maintains rich system catalogs (e.g., sys.columns, sys.tables, sys.foreign_keys) and the standardized INFORMATION_SCHEMA views (e.g., INFORMATION_SCHEMA.COLUMNS, INFORMATION_SCHEMA.PARAMETERS), which contain comprehensive information about the structure of the objects. This metadata provides the necessary context (data types, dimensions, field nullability), which allows the proposed abstraction layer to algorithmically generate valid test scenarios without relying on predefined static values.

3. Proposed Metadata-Based Approach

The proposed metadata-based model builds upon the core functionality of tSQLt by encapsulating the test preparation and writing process through the introduction of three abstract, interdependent layers. This approach ensures loose coupling and is specifically engineered for database-centric architectures where configurations and schemas are distributed (such as in multi-tenant systems or client-based segmentation) (Fig. 1):

- Metadata Extraction Layer

This layer functions as the analytical module of the system. Through static code analysis of the target stored procedure (utilizing system functions such as sys.dm_sql_referenced_entities), the layer automatically constructs a dependency graph, discovering all bound tables and views. Based on this graph, queries are dispatched to the system catalogs to extract critical technical parameters for each attribute: data types, dimensions, and nullability flags (IS_NULLABLE).

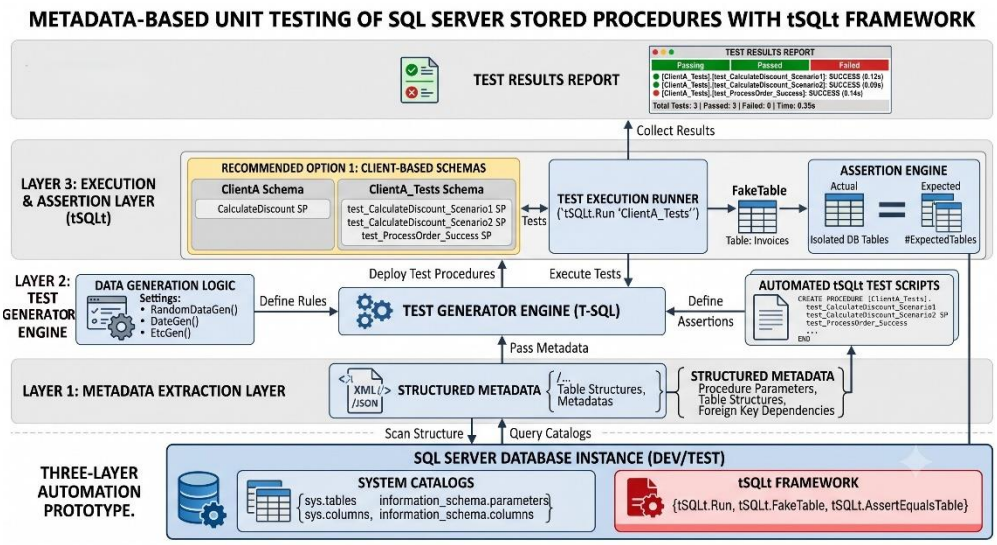


Fig. 1. Three-Layer Metadata-Driven Architecture Pattern

- Test Generator Engine

This layer represents the computational core of the prototype, implemented entirely via T-SQL stored procedures. The process of automatically generating consistent test data within this engine shares numerous common challenges with modern data preparation methods (Data Preparation). As noted in research regarding data preparation for analytical purposes and machine learning, automating these steps is of key importance for increasing the speed, reliability, and accuracy of final outcomes [5].

Applying the same philosophy, the proposed engine completely eliminates manual labor by receiving the structured metadata from the first abstraction layer and applying deterministic production rules for alignment (mapping). Through complex dynamic SQL (Dynamic SQL) generation algorithms, the engine synthesizes ready and valid T-SQL code for the respective tSQLt test.

To optimize the computational overhead during execution, the engine automatically generates minimal baseline values according to the discovered data types in the catalogs (e.g., sequential incremental numbers for INT, current system time for DATETIME, and standardized text markers for VARCHAR). A vital architectural characteristic is that by deliberately omitting all columns that permit missing values (NULL), the generator creates a so-called "minimal semantic intersection" of the test data. This protects the test environment from being overloaded with redundant informational ballast and radically conserves DBMS computational resources during mass iterations.

- Execution & Assertion Layer

This layer manages the lifecycle and deployment of tests by invoking tSQLt.Run. Its primary role is to orchestrate transactional isolation: it replaces real tables with their virtual copies (via tSQLt.FakeTable), injects the data automatically synthesized by the previous layer, and executes the tested stored procedure.

During the verification phase (Assertion), the layer addresses one of the fundamental dilemmas in software testing theory - the so-called Oracle Problem, which manifests as the difficulty of defining an automated mechanism that flawlessly distinguishes correct from incorrect system behavior [6]. In relational databases, this problem is highly pronounced due

to the multidimensionality of data. The proposed abstraction layer solves this dilemma elegantly through bulk relational verification. Instead of vulnerable scalar checks, the final state of objects is compared cell by cell using tSQLt.AssertEqualsTable. This function acts as a deterministic tabular oracle that validates the comprehensive structure and generates a detailed differential report upon the slightest discrepancy.

4. Business Logic Abstraction via Configuration Meta-Tables

System catalogs (such as INFORMATION_SCHEMA) do not possess semantic knowledge regarding the specific business domain. For example, a column named [CountryCode] with a data type of VARCHAR(2) would automatically receive a default generated value of 'Au' from the baseline engine. This could trigger a runtime error if the underlying business logic strictly requires a valid ISO country code. An identical problem occurs with rigidly formatted data such as IBANs, national identification numbers, or specific nomenclature keys.

To transition the prototype into a fully capable enterprise solution, an supplementary relational metadata structure has been integrated. This structure enables centralized management of specific business rules that system catalogs cannot inherently anticipate:

- AutomationEngine.ColumnRules: Functions as a global dictionary allowing the definition of static overrides. For example: "If the attribute name is StatusID, always generate a value of 1." The generation engine validates the name of each column against this dictionary before applying the baseline function $f(a)$.
- AutomationEngine.ProcedureScenarios: Enables the definition of multiple test scenarios for specific stored procedures, thereby facilitating boundary value testing (Boundary Value Testing). Different hardcoded values can be defined for a single procedure (e.g., OrderAmount = 0 versus OrderAmount = 10000). The engine extracts this metadata and algorithmically synthesizes independent test methods that cover the various execution branches (branches) of the code.

5. Schema Design and Scale Management

Managing massive arrays of objects in relational databases introduces architectural challenges similar to those found in organizing big data repositories and data lakes [7]. To mitigate structural chaos (Object Explosion) as the testing infrastructure scales, the proposed model introduces a specialized topology.

Standard practices within the tSQLt framework recommend creating a separate schema (Test Class) for each tested functionality. However, in environments with hundreds of stored procedures, this approach triggers an object explosion (Object Explosion) within the database. Consequently, this heavily complicates navigation and administration inside management suites such as SQL Server Management Studio (SSMS).

Since application-level databases are frequently structured logically using schemas that correspond to specific clients (e.g., [ClientA], [ClientB]) alongside common features in the [dbo] schema, the model introduces a Consolidated Schema Approach (Zonal / Client-based Partitioning) [8].

Instead of maintaining hundreds of individual test classes, exactly one test schema is defined per business zone or client (e.g., [ClientA_Tests]). All automatically generated test procedures are assigned to this single schema, adhering to a strict naming convention: [TestSchemaName].[test_OriginalProcedureName_AutomatedScenario].

Object lifecycle management during deployment (Deployment) is strictly enforced through three distinct isolation levels:

- [tSQLt]: Shared infrastructural library. It contains system assertions and core isolation logic. It is entirely omitted from production deployment scripts via schema filters.
- [Client_Tests]: Container for automated client-specific tests. It aggregates hundreds of test methods into a single schema. This container is dropped automatically or skipped during compilation to the live production server.
- [ClientA] / [dbo]: Core business layers containing the actual code and tables subject to verification. This is the only component that is physically deployed onto the production server.

6. Practical Example and Code Generation (Case Study)

To demonstrate the real-world applicability of the model, this section presents a practical case study inspired by the computational challenges within the Intelligent Risk Management for Businesses from Adverse Events and Natural Disasters project. In such risk-evaluation environments, it is parameter-critical to validate system behavior when specific operational or hazard thresholds (boundary values) are crossed. For the sake of structural clarity, this transformation process is illustrated through a generalized business procedure, dbo.CalculateOrderDiscount, which evaluates thresholds based on two related tables: Customers and Orders, linked via a strict foreign key relationship (Foreign Key).

- Original Schema of the Database Tables

```
CREATE TABLE dbo.Customers (
    CustomerID INT IDENTITY(1,1) PRIMARY KEY,
    CustomerName NVARCHAR(100) NOT NULL,
    IsPremium BIT NOT NULL -- Flag for VIP customers
);

CREATE TABLE dbo.Orders (
    OrderID INT PRIMARY KEY,
    CustomerID INT NOT NULL FOREIGN KEY REFERENCES
dbo.Customers(CustomerID),
    OrderAmount DECIMAL(18,2) NOT NULL,
    OrderDate DATETIME NOT NULL,
    ShippingNotes NVARCHAR(MAX) NULL
);
```

- Defining the Test Scenario in the Metadata

Instead of merely testing a successful execution, the goal is to verify the business rule: "If the order exceeds 10000 euro, the customer receives a discount." To achieve this, the following condition is predefined in the system configuration table AutomationEngine.ProcedureScenarios:

Target Procedure: CalculateOrderDiscount
Override Column: OrderAmount
Value: 10000.00

- Test Code Synthesis

Upon starting the generator, Layer 1 scans the procedure and discovers its dependency on both tables. It analyzes their metadata, identifies the foreign key, and extracts the data types.

Subsequently, Layer 2 reads the configuration for the boundary scenario and compiles the following fully completed tSQLt script:

```
CREATE PROCEDURE
[ClientA_Tests].[test_CalculateOrderDiscount_HighValueScenario]
  AS
  BEGIN
      -- 1. TRANSACTIONAL ISOLATION AND REMOVAL OF FOREIGN KEYS
      EXEC tSQLt.FakeTable 'dbo.Customers';
      EXEC tSQLt.FakeTable 'dbo.Orders';
      -- 2. DATA SYNTHESIS (SETUP)
      -- Generating mandatory fields for Customer
      INSERT INTO dbo.Customers (CustomerID, CustomerName, IsPremium)
      VALUES (1, 'Auto_CustomerName', 0);
      -- Generating Order: Foreign Key is logically respected, and
  the amount is overridden by the meta-table
      INSERT INTO dbo.Orders (OrderID, CustomerID, OrderAmount,
  OrderDate)
      VALUES (1, 1, 10000.00, GETDATE());
      -- 3. BUSINESS LOGICAL EXECUTION (ACT)
      -- EXEC dbo.CalculateOrderDiscount @OrderID = 1;
      -- 4. VERIFICATION (ASSERT)
      -- EXEC tSQLt.AssertEqualsTable '#Expected', 'Actual';
  END;
```

- Analysis of the Results:

Decoupling of Relations: By executing tSQLt.FakeTable sequentially on both objects, the model completely eliminates the physical foreign key constraint binding Orders and Customers. Furthermore, it strips the IDENTITY constraint specification, enabling direct value injection of CustomerID = 1.

Semantic Optimization: As designed in the baseline engine, the optional column ShippingNotes is automatically omitted to achieve a minimal semantic intersection of the test transaction data.

Business Intelligence in Action: While the standard engine would typically fallback to generating a dummy default value of 100.00 for a DECIMAL data type, the integration of configuration metadata allows the system to intelligently override it with 10000.00. This creates a perfectly valid boundary test scenario without requiring a single row of manually written code.

This automated execution finishes in fractions of a second, entirely eliminating human error and removing the overhead of manually synchronizing test scripts during database schema modifications (Schema Drift).

7. Conclusion

The developed metadata-based model and software engine built on top of the tSQLt framework offer a deterministic and highly scalable solution for unit testing stored procedures. Experimental results from deploying the model in a real development environment (Dev Environment) containing over 200 stored procedures demonstrate a reduction in test creation time of more than 70%.

The most significant qualitative outcome is the complete elimination of manual labor required to synchronize test scripts during database schema modifications (Schema Drift) [9]. The integration of transactional control and automated metadata extraction proves that quality

assurance in relational databases can be managed using the same modern principles applied to application software.

Future directions for this research include the integration of Artificial Intelligence (AI) and Natural Language Processing (NLP) models to perform deep semantic analysis of the T-SQL code itself and its accompanying documentation. Adapting modern methodological approaches for context extraction and classification of unstructured data-similar to those applied in multimodal sentiment analysis [10] - would enable the automated prediction and definition of test scenarios for business logic. Such a symbiosis between database automation and innovative analytical information technologies would elevate the process of digitalization and quality control in the digital economy to the next level.

Acknowledgements

This work has been carried out in the framework of the Project "Smart Risk Management for Business from Adverse Events and Natural Disasters", supported by the National Science Fund of Bulgaria, Grand № KP-06-H55/5 from 15.11.2021.

References

1. Ambler, S. W. (2003). Agile Database Techniques: Effective Strategies for the Agile Software Developer. John Wiley & Sons.
2. Sadalage, P. J., & Fowler, M. (2012). NoSQL Distilled: A Brief Guide to the Emerging World of Polyglot Persistence. Addison-Wesley.
3. tSQLt Team. (2025). tSQLt - Database Unit Testing Framework for SQL Server. Official Documentation available at https://tsqlt.org.
4. Harrison, G. (2023). Next Generation Databases: NoSQL, NewSQL, and Big Data. Apress.
5. Miteva, G., & Murdzheva, A. (2023). Data Preparation Techniques And Platforms In The Context Of Machine Learning. Proceedings of the Scientific Conference „Innovative Information Technologies for Economy Digitalization" (IITED), UNWE, Sofia, Bulgaria.
6. Barr, E. T., Harman, M., McMinn, P., Shahbaz, M., & Yoo, S. (2015). The oracle problem in software testing: A survey. IEEE transactions on software engineering, 41(5), 507-525.
7. Stefanov, G. (2023). Data lakes as data store for big data. Analyzing AWS approach. Innovative Information Technologies for Economy Digitalization (IITED) 2023.
8. Celko, J. (2005). Joe Celko's SQL Programming Style. Morgan Kaufmann.
9. Fowler, M. (2018). Refactoring: Improving the Design of Existing Code (2nd Edition). Addison-Wesley.
10. Yordanova, S. (2025). Methodological Approaches in Multimodal Sentiment Analysis. 3rd Scientific Conference „Innovative Information Technologies for Economy Digitalization" (IITED – 2025), October 2025, UNWE, Sofia, Bulgaria.

AI Aiding Financial Control

Milen Markov

University of National and World Economy, Sofia, Bulgaria
milen.markov@unwe.bg

Abstract. Global financial transactions are becoming increasingly digital. This process has made traditional financial control, predicated upon historical data analysis and statistical data sampling, outdated. This paper examines the paradigm shift in financial control toward introducing Artificial Intelligence (AI) in corporate and public financial oversight. Integrating Machine Learning (ML), Natural Language Processing (NLP), Robotic Process Automation (RPA), and Digital Twins facilitates a continuous auditing across entire transactional populations rather than only samples of them. Automated AI tools and models enhance detection of fraudulent activities, mitigate business and operational risks, counteract the shadow economy, and diminish fiscal deficits. A case-in-point is the shrinking European Union (EU) Value Added Tax (VAT) gap, which followed as a result of mandatory electronic invoicing across the EU. AI also introduces critical challenges for users of transparency of algorithms, and of compliance with legal frameworks. The author argues that AI will not displace the need to employ financial control professionals, but transform the requirements to the job by making sure people, who enter into this profession, have a much deeper understanding of the different AI tools and models. In the future financial governance shall work as a symbiosis between a highly trained human, who will provide the contextual meaning, and a machine, which has the compute power to go through an enormous amount of structured and unstructured data.

Keywords. Financial Control, Artificial Intelligence, Continuous Auditing, Tax Compliance, Machine Learning,

1. Introduction

The world is going hyper-digital. Financial transactions in all areas of business are growing exponentially. A surge in the velocity, variety and volume of transactions is evident in practically all countries around the world. The shift towards more and more digital consumption is irreversible. Stimulated by the Covid-19 epidemic e-commerce sales in the European Union (EU) alone rose from 14.9% of total retail volume in 2019 to 21.8% by 2023 [1].

The fast increase in the volume of transactions has introduced new challenges to financial oversight. Only until a couple of years ago, the correct application of the frameworks for financial control was predicated upon historical data analysis and statistical random sampling. However, these traditional methods for financial monitoring have become outdated.

The global shifts in consumption patterns, accompanied by intense geopolitical crises (such as raging wars, uncontrollable immigration, soaring energy prices etc.), create a seriously unstable political and economic environment. In these uncertain times maintaining macroeconomic stability requires financial governance and control, which can efficiently deal with the new challenges.

The stakes for governments are high. Sound financial control over the most important budgetary revenue streams becomes of key importance. In Bulgaria for instance, the Value

Added Tax (VAT) generates the largest share of state budget revenues, while the existing shadow economy remains a significant challenge.

A pillar of corporate management and public governance, financial control plays an ever more important part. It is devised to confirm the empirical correctness of financial reports, mitigate financial risks, reduce vulnerabilities and exposures, and compel compliance with national, regional, local and company-wide regulatory frameworks across all economic activities.

2. Traditional audit methodologies

Using traditional financial control methodologies such as analysis of historical data and random statistical sampling, applied widely by professionals in the field, it is typically possible to evaluate only an average of 2% to 5% of a given transactional population [2, 3]. While economically justified in the era of analog reporting and analysis, this limited sampling is fundamentally incomplete. It leaves out massive amounts of data unexamined. Hence, it is inherently impossible to cover financial datasets in their entirety in order to analyze them in depth and breadth. This vulnerability provides fertile ground for micro frauds, because systemic inefficiencies, more often than not, go undetected.

The possibilities for financial oversight and their efficacy increased dramatically after the deployment of Artificial Intelligence (AI). Big data architecture, which allows advanced analytics and predictive modelling, paved the way for an evolutionary leap - a paradigm shift in the methodologies used to sample data and reflect on the findings [4, 5]. By using high compute power to go through, package and analyze massive and varied datasets, AI is capable to go beyond the limitations of sample-based auditing [2].

The aid, which AI started providing to professionals in the financial control field is transformative. It allows the uninterrupted, dynamic monitoring of entire financial populations, thereby transitioning financial control from a reactive forensic approach to a proactive, predictive discipline [6].

Introducing modern digital technologies and software into financial control is not only intended to help prevent fraud, it is also tied to building a more sustainable corporate culture and to guaranteeing efficient, streamlined and lean business processes. Corporate sustainability is focused on economic, social, and environmental performance. Therefore, the modern financial controller role is to use dynamic data sets to make sure that an organization remains sustainable across all three fields [7].

Modelling the business processes in an organization as an independent mathematical system could help assess the economic perspectives and stability of corporations with complex hierarchical structures. By applying the Lyapunov's function financial controllers and analysts can construct a stress score, with the help of which they can assess the actual business conditions and compare them with the planned target KPIs of a company, thus identifying budget or time overruns, diminishing product quality, excess debt etc.

Controllers can then put in place policies, measures, rules and safeguards (such as spending restrictions, debt limits, floors and caps on capital funds) to continuously ensure that the stress score is in the predefined corridor and external or internal disruptions do not turn into business blunders. Gathering information digitally, keeping big data sets and using AI tools for continuous dynamic evaluation is critical in this endeavor.

3. Machine Learning

An important branch of AI is Machine Learning (ML). Used to identify patterns in historical data and make decisions without being explicitly programmed for that, some ML

algorithms are capable of detecting fraudulent activities with very high precision [8]. ML is also successfully used in tax control for performing tax compliance on a national and/or federal level.

The frameworks used traditionally for fraud detection are largely based on rules, protocols and regulatory frameworks. They rely heavily in static data and often generate false-positive rates of 80% to 90% [3]. Controlled algorithms and neural networks, trained by using the methods of MS on historical sets of verified data a far superior. They are capable of predicting tax evasion, money laundering, concealment of assets and/or income as well as other illicit actions much easier.

Integrating ML in the process of financial statement audits removes the limitations of the traditional models and allows for a multivariate analysis of the financial health of an organization [5, 9]. In addition, at macro level digital models for financial control carry the benefits of increased revenue and streamlined revenue service operations. Proof for that is the significant decrease in the average VAT revenue gap for the EU-27 observed from 2018 to 2022. In this period the revenue gap dropped from 11.1% to 6.9%. This improvement is directly connected to the introduction by member-states in their tax systems of modern digital control mechanisms and dynamic real-time reporting of financial transactions and of asset procurement/liquidation [1].

According to the learned literature micro and small enterprises often manipulate financial reporting and/or sales in order to stay below VAT registration limits. AI and ML algorithms surpass any other method in finding and discovering statistical clusters of sales, induced artificially by companies in order to skip or postpone VAT registration [10].

Automating financial reporting with the help of AI has significantly modernized the processes for ensuring compliance. EU digital control instruments, such as the One-Stop-Shop regime for e-commerce (OSS), rely heavily on continuous monitoring based on mathematical algorithms.

Another important area of application of ML is the financial control of the mandatory electronic invoicing across B2B transactioncs in the EU, which has to be implemented under the "VAT in the Digital Age" - the EU ViDA initiative. In countries such as Portugal and Italy, where mandatory e-invoicing was introduced back in 2019 and 2014, respectively, ML auditing is one of the main contributors to closing the VAT gap. In addition, crises induced tax stimuli such as temporary VAT cuts and/or tax rate modifications (including Pigouvian sin taxes) rely heavily on automated AI systems [11].

4. Natural Language Processing

Around 80% of modern data collected by companies and governments are unstructured and they cannot be analyzed by ML. This data is contextual and included in a big array of sources such as contracts, company communications, regulatory filings, public announcements etc. AI models, that could be used to analyze this wealth of information, rely on Natural Language Processing (NLP). These models provide digital compute power and specific architecture, required to assess this enormous data ocean [8, 12].

Models, which use NLP, apply deep learning algorithms to digest, identify, analyze and parametrize critical KPIs. They can do it with almost human precision. In regulatory compliance and contract management, NLP models can be invaluable. They are already used by internal auditors and compliance officers to review tens of thousands of contract agreements and juxtapose them to vendor shipments, payments, quality of goods produced and received etc. They can also be used for streamlining operations, ensuring strict adherence to contractual obligations and to the international accounting standards used for reporting them.

Using big data sets of unstructured data as defensible audit trails changes significantly the evidence collecting process in financial and compliance audits and enhances essentially the process of risk assessment [2].

With the help of Robotic Process Automation (RPA), used as the operational layer by some AI models, a number of real-time insights can be collected [7]. RPA functions by applying programmable software bots to perform repetitive, rule-based work processes across different corporate systems. In financial control it could be used to reconcile and automate complex numerical and contextual data. When integrated with ML it is called "Intelligent Automation". Such models are capable of not only reconciling discrepancies in low priority items (such as work schedules), but also automatically recognizing statistically significant deviations and directing information about them as well as proposing possible solutions to human internal financial auditors and controllers [6].

This process, called Continuous Auditing (CA), gives organizations the tools to analyze the entirety of their transactional population, thus dramatically minimizing the risk of overlooking vulnerabilities in the audit ecosystem. It transforms the auditing procedures from a sample-based modelling to a dynamic real-time assurance [4, 6].

5. Digital Twins

A digital copy of a physical entity or a process, digital twins are used by financial auditors for dynamic monitoring, analysis and for virtual simulation of possible optimization steps [13, 14]. If the performance of the digital twin improves after introducing the updates, then the findings are used as basis for improving the physical entity or process as well. Digital twins are created with the help of all AI tools discussed above - ML, NLP, and RPA.

In financial control developing a digital twin of a company's Enterprise Resource Planning (ERP) system allows internal controllers and auditors to identify potential financial risks and vulnerabilities, to simulate financial stress tests and to follow the general financial well-being of the company in different contextual scenarios, including both scenarios with shift in the impact on the company of external factors (such as increased customer expectations on quality) and of internal factors (such as company-wide compliance with work place safety).

Financial controllers and auditors use digital twins in many different industries including, but not limited to transportation, manufacturing, energy production and supply, healthcare, education, and defense. [13]. Digital avatars are utilized to increase efficiency and efficacy, to improve customer service, to train staff, to improve overall product satisfaction etc. [14]. The methodology for creating such digital avatars involves high-tech 3D scanning and advanced motion capture systems.

More and more often digital twins in the financial sector serve as AI independent financial advisors or compliance officers, who engage internal and external stakeholders to enforce financial policies in real-time. Such online digital advisors take care of mundane everyday tasks like offering and selling standardized financial products and services, helping customers with their financial transactions, and even providing legal and business advice on acquiring new stocks, bonds, insurance and assets.

6. Ethical concerns with AI

The advantages of using AI tools in financial control are immense. The empirical evidence is overwhelming. However, deploying these tools in practice requires a highly skilled work force and profound ethical and regulatory frameworks [9, 14].

Without financial controllers who are well-versed in both applicable accounting standards and regulatory compliance, as well as data sets hierarchy and big data modelling, AI

algorithms risk turning into useless and expensive tools, which only increase uncertainty of risk events rather then remove it. Hence, companies and governments have to perform continuous cutting-edge training of their employees in order to enable the workforce to use the AI tools as intended and extract the highest possible value out of the human-machine interaction for the shareholders, the customers and the general public.

The ethical concerns of creating and using AI tools and digital human avatars have to be taken into account as well. Considerations about identity, human rights, and responsibility are mentioned by some authors as quite important in the discourse when it comes to AI [14].

A very important worry concerns the autonomy, which AI tools can have. If they are taking a decision on their own and act independently to greenlight or ban a transaction, who is responsible if the decision was wrong - is it the human on which account the decision was taken, the producer of the AI tool, representatives of the organization, which employed the AI tool (private or public company, government agency), the regulator, who allowed the tool to be used, or is it someone else?

Another important issue arises from the concern about ownership and control - who owns the financial data and who is responsible for decisions taken based on these data by a digital avatar.

Modern AI models often function as sealed airplane black boxes. The algorithm of their operations is not known to the owner; it is only known to the producer. This violates auditing standards for transparency, because there is no confirmable and comprehensive evidence [2]. Financial controllers must be able to explain the AI tools and models used to evaluate and mitigate risk to all sorts of stakeholders including regulators.

Further ethical concerns include the need for AI tools and models to make sure there is integrated security and abuse prevention in them. If hacked digital twins and AI control algorithms may be used maliciously to harm the person, who was trying to detect fraud and vulnerabilities in the system in the first place, or for corporate espionage of the entity, which ordered the audit.

Additionally, instilling trust in users in AI tools and models hinges on their transparency and openness regarding informing the human that they are interacting with an artificial intelligence rather than with another human being.

7. The Future

Integrating AI tools and building AI models to perform important analytical financial control procedures raises the question about virtual personas taking away jobs from real people - highly trained employees. Analyses of recent labor market developments, however, show that AI aids and augments significantly the work of humans in the financial sector, but does not displace them [15, 16].

Nevertheless, modern qualifications of a skilled financial controller and/or auditor have to include being proficient in data science, and understanding algorithms. To some extent these qualifications become more and more important than traditional accounting knowledge and expertise [15].

Finally, a human has to remain an integral part of the decision-making process, because AI tools and models are devised to look at probabilities. They are not very good when it comes down to ethical concerns or contextual details. Therefore, professional human skepticism and human contextual judgment, based on gathered experience, are of the highest importance when final decisions are being made [8, 12]. Taking into account these ethical concerns, companies and governments should ensure, that AI tools and models they deploy align with societal rules and values [13, 14].

This shows that the future of financial control belongs to a new way of working. A new way constituting a cooperative synergy between AI, which can deliver the computational power and precision, and highly trained human individuals, who are capable of providing strategic insight, by taking into account ethical and moral concerns, thus providing much needed contextual guidance for the AI. A paradigm shift in financial control is afoot and it is evolving. Technological advancements together with human supervision must work together in order to maintain trust and regulatory compliance.

8. Conclusion

By integrating AI into financial control we are poised to accomplish a quantitative leap in financial and compliance auditing, and, through that, in corporate and public governance. AI tools and models do not have the same limitations as traditional approaches for analyses do. Before the introduction of AI financial controllers and auditors relied heavily on historical data and statistical sampling.

Technologies such as Machine Learning, Natural Language Processing, and Robotic Process Automation facilitate uninterrupted, dynamic, real-time monitoring of entire financial populations, thus reducing vulnerabilities and guaranteeing a level of assessment unseen and impossible before.

Empirical data suggests that AI tools can be used to remove fiscal hemorrhages and help devise a more sustainable resource allocation. In addition, AI can help to streamline operations, counteract the grey economy and generate more tax income. Connecting fiscal devices with national revenue services is a case-in-point.

The modernization of the financial control through AI tools and models requires transparency of the used algorithms, maintaining data privacy and fighting data theft, as well as the ability to manage virtual identities. However, human oversight, and relying on ethical judgement, will remain needed [16]. Advanced autonomous models can ensure the organizational goals could be closely aligned with both societal values and with comprehensive regulatory frameworks.

References

1. Georgieva, P. (2022). VAT Compliance and Digital Economy: Challenges for Bulgarian Tax Administration. Economic Alternatives, (3), 412-428.
2. Brown-Liburd, H., & Vasarhelyi, M. A. (2015). Big Data and Audit Evidence. Journal of Emerging Technologies in Accounting, 12(1), 1-16.
3. Earley, C. E. (2015). Data analytics in auditing: Opportunities and challenges. Business Horizons, 58(5), 493-500.
4. Appelbaum, D., Kogan, A., & Vasarhelyi, M. A. (2017). Big Data and Analytics in the Modern Audit Engagement: Research Needs. Auditing: A Journal of Practice & Theory, 36(4), 1-27.
5. Cao, M., Chychyla, R., & Stewart, T. (2015). Big Data Analytics in Financial Statement Audits. Accounting Horizons, 29(2), 423-429.
6. Zhang, J., Yang, X., & Appelbaum, D. (2015). Toward Effective Big Data Analysis in Continuous Auditing. Accounting Horizons, 29(2), 469-476.
7. Kokina, J., & Davenport, T. H. (2017). The Emergence of Artificial Intelligence: How Automation is Changing Auditing. Journal of Emerging Technologies in Accounting, 14(1), 115-122.
8. Davenport, T. H., & Ronanki, R. (2018). Artificial Intelligence for the Real World. Harvard Business Review, 96(1), 108-116.
9. Krahel, J. P., & Titera, W. N. (2015). Consequences of Big Data and Formalization on Accounting and Auditing Standards. Accounting Horizons, 29(2), 409-422.
10. Weichenrieder, A.J., & Ehrenberg, A. (2020). Threshold effects in VAT registration: evidence from Germany. Fiscal Studies, 41(3), 547-571.

11. Fuest, C., Neumeier, F., & Stimmelmayr, M. (2021). VAT pass-through: the case of Germany's temporary VAT cut. CESifo Working Paper No. 9120.
12. Bizarro, P. A., & Dorfman, M. (2017). Artificial Intelligence: The Future of Auditing. Internal Auditing, 32(5), 21-26.
13. Bennett, D., Vasileva, V., & Aleksandrova, V. (2024). The Integration of State-of-the-Art Technologies and the Ethical Considerations in the Development of Digital Twins. Selected Papers from the 14th ICAICTSEE, UNWE.
14. Helbing, D., & Argota Sanchez-Vaquerizo, J. (2022). Digital Twins: Potentials, Ethical Issues, and Limitations. Handbook on the Politics and Governance of Big Data and Artificial Intelligence, Edward Elgar.
15. Issa, H., Sun, T., & Vasarhelyi, M. A. (2016). Research Ideas for Artificial Intelligence in Auditing: The Formalization of Audit and Workforce Supplementation. Journal of Emerging Technologies in Accounting, 13(2), 1-20.
16. Sutton, S. G., Holt, M., & Arnold, V. (2016). The reports of my death are greatly exaggerated - Artificial intelligence research in accounting. International Journal of Accounting Information Systems, 22, 60-73.

A Model-Driven Approach to Information Security of Automated and Autonomous Systems in Industry 4.0 Environments

Ivan Gaidarski

Laboratory of Unmanned Robotics Systems,
Institute of Robotics, Bulgarian Academy of Sciences, Sofia, Bulgaria
ivangaidarski@ir.bas.bg

Abstract. The article presents a model-driven approach to designing information security for automated and autonomous systems within an Industry 4.0 environment. The approach adapts a methodology for developing information security systems based on top-down design, conceptual modeling, and model transformation into an object-oriented UML design model. A framework is proposed that links the autonomous system's boundaries, risk analysis, its information and physical assets, and protective mechanisms. A generalized six-component conceptual model is adapted, representing the object to be protected via the "sensor – processing – decision – command – execution" control loop. The architecture and functionality are described using a UML class diagram and a sequence diagram. A generalized quantitative risk assessment model is proposed and applied to an illustrative scenario involving an autonomous mobile robot, resulting in a reduction of residual system risk by approximately 61%. This approach can serve as a foundation for more detailed models, simulations, and the implementation of security functions in Industry 4.0 automated and autonomous systems.

Keywords. Information security, automated systems, autonomous systems, Industry 4.0, model-driven approach.

1. Introduction

Industry 4.0 is an environment that integrates traditional industrial manufacturing facilities with information and communication technologies, the Internet of Things (IoT), the Internet of Services (IoS), and automated control systems. One of the concepts associated with Industry 4.0 is Cyber-Physical Systems (CPS). These are systems that integrate sensing, control, communication, computing, and physical processes.

In the case of autonomous systems, mechanisms for perception, data processing, decision-making and physical interaction with the environment are added to this environment. This integration creates a direct link between the information and physical aspects of the industrial system. Information security breaches are not limited only to data leakage or loss. They can lead to serious consequences, such as incorrect decisions, functional failures, and industrial accidents.

Designing Industry 4.0 systems is a complex process involving the convergence of operational and information technologies, as well as the combination of various technical approaches, protocols, and security technologies. This complex mix of technologies complicates the design process and makes it difficult to trace the relationships between system requirements, architecture, and implementation.

It is crucial that information security is not viewed as an additional layer but rather as an integral part of the industrial system design process. In previous studies, the author proposed

a model-based, top-down approach for developing information security systems (ISS), involving a sequential progression through problem domain analysis, a conceptual model, a design model, and an implementation model [2]. The approach is technology-independent and employs "model-to-model" transformation; within the design model, architecture and functionality are described using the object-oriented modeling language UML.

In this article, the approach is adapted to address the information security of automated and autonomous systems within Industry 4.0. The approach entails constructing a unified model that links system information security requirements with assets, threats, architectural components, and protective measures from the early stages of design.

2. Specifics of Information Security in Automated and Autonomous Systems

Automated systems operate by executing pre defined functions and rules. In contrast, autonomous systems are capable of perceiving information from the environment, processing that information, and independently deciding upon and executing the necessary actions. For the purposes of information security, both types of systems can be represented as a combination of sensing, communication, computational, control, and actuation components.

Information security is described using key concepts such as asset, vulnerability, threat, threat source, attack, incident, and countermeasure [2]. This terminology applies to both automated and autonomous systems. The difference lies in the fact that assets include not only data and IT resources but also sensors, controllers, communication interfaces, decision-making models, and physical process parameters.

For analysis purposes, the system can be divided into functional areas: perception (sensors), communication, decision-making with edge processing and AI, control, actuation, and operator interaction. Each area is characterized by specific vulnerabilities as well as certain interdependencies between them. For instance, spoofed sensor data can lead to an incorrect decision by the AI module, which subsequently results in wrong command to the actuators (Table 1).

Table 1. Assets and typical threats in an automated and autonomous system

Area	Main assets	Typical threats	Main measures
Sensors	Measurements, identifiers, configurations	Data spoofing or injection	Validation, authentication, monitoring
Communications	Messages, control signals	Eavesdropping, tampering, denial of service (DoS)	Encryption, segmentation, access control
Edge and AI	Models, code, parameters, data	Malware, model manipulation	Protected storage, version control, integrity checks
Control	Rules, configurations, commands	Unauthorized modification	AAA, logging, multi-factor authentication
Actuation	Commands and physical actions	Wrong command	Command checks, safe state, limits
Operator	Identity, privileges, data	Misuse, social engineering	Role-based access control, training

Consequently, protecting individual computers or network segments is insufficient. It is necessary to analyze the "asset–vulnerability–threat–attack–impact" chain and track how a breach propagates across logical and physical components.

In Industry 4.0, the attack surface expands due to the convergence of IT and Operational Technology (OT), sensors and other IoT devices, edge computing, remote maintenance, and integration with various enterprise and cloud services [3,4,5].

Security requirements must encompass not only the elements of the CIA triad—Confidentiality, Integrity, and Availability, but also controlled management, process continuity, and the transition to a safe state—an area where security intersects with functional safety [6].

3. Model-Driven Approach to Developing the Security System

The proposed approach retains the core logic of the Model-Driven Approach for Designing of Information Security System [2]:

1. Defining a framework for describing the architecture and viewpoints [7],
2. Analyzing the problem domain,
3. Constructing a conceptual model,
4. Transforming the conceptual model into a design model,
5. Constructing a simulation model and an implementation model.

When applying the method to an autonomous system, the sequence remains the same, but the content of the models is expanded to include additional elements: physical processes, sensor data, AI components, and control functions.

The adapted method comprises five stages (Fig. 1): defining system boundaries and stakeholders; analyzing assets, threats, vulnerabilities, and risks; constructing a conceptual model; transforming it into a UML design model; and selecting an implementation environment and verifying the model.

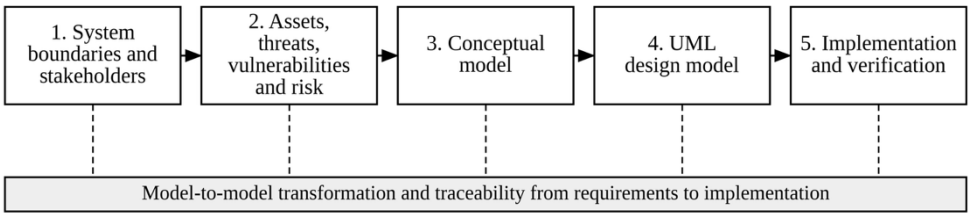


Fig. 1. Adapted model-driven process for developing information security

Stakeholders include company management, legal experts, design engineers, process engineers, production automation specialists, software developers, IT specialists, information security experts, system administrators, operators, product specialists, and others. Their perspectives are consolidated into a unified model representing the requirements for the industrial system, thereby reducing the risk of overlooking requirements or encountering conflicts between them. The primary advantage of this approach is the system's platform independence, achieved by decoupling the architecture from specific technologies. The same conceptual model can be implemented using various technological components, such as controllers, communication protocols, edge platforms, or AI modules.

4. Conceptual model of information security

The generalized conceptual model of the information security system (ISS) [2] includes six components: “Security Model and Policy”, “Management and Configuration”, “Monitoring and Analysis”, “Endpoint Protection”, “Communication Protection” and “Data

Protection" (Fig. 2). A "Security Model and Policy" defines what is protected and why, and with what restrictions. In the case of an autonomous system, the security policy, in addition to traditional concepts, also includes rules for permissible actions, permissible ranges of measured values, critical commands and production safety conditions. "Management and configuration", configures the protection of endpoints (sensor nodes, controllers, edge computers, operator stations), of communications (industrial networks and wireless channels) and of data (at rest, in motion and in use) depending on the adopted security policy. "Monitoring and Analysis" collects data on events (logs), alarms and deviation indicators and returns feedback on the state of the system.

The significant change from the organizational ISS model [2] is that the protected object includes the physical environment. The result of information processing can be related to certain physical actions, therefore the model tracks, in addition to the information flow, the control loop "sensor - processing - decision - command - executive mechanism".

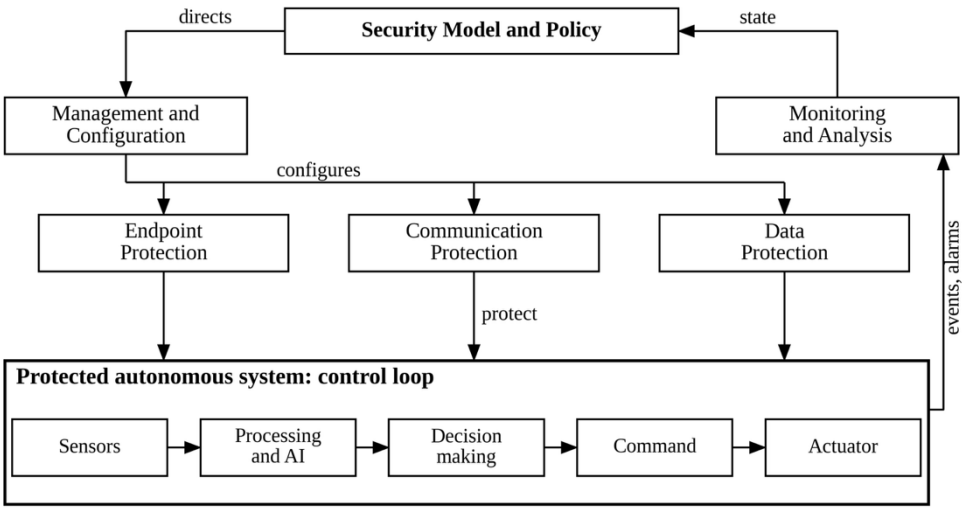


Fig. 2. Adapted generalized conceptual model of the information security of an autonomous system

5. UML Design Model

The next stage involves transforming the conceptual model into an object-oriented design model. The resulting model is examined from both architectural and functional perspectives [2]. Using UML notation [8,9], the static structure of the conceptual model is represented by a class diagram, while the interaction during a critical scenario is depicted via a sequence diagram. The class diagram (Fig. 3) comprises the core logical objects of the secure autonomous system: SensorNode, EdgeAI, DecisionControl, Actuator, SecurityPolicy, and MonitoringAnalysis. The control loop connections ("data," "result," "command") describe normal operation, whereas the "use" dependencies pointing to SecurityPolicy represent the checks performed by EdgeAI (check) and DecisionControl (authorize) prior to transmitting a result or command. MonitoringAnalysis receives information regarding the system's status and security policy violations. The model is technology independent and is not tied to specific hardware or software solutions, such as controllers, protocols, operating systems or communication technologies.

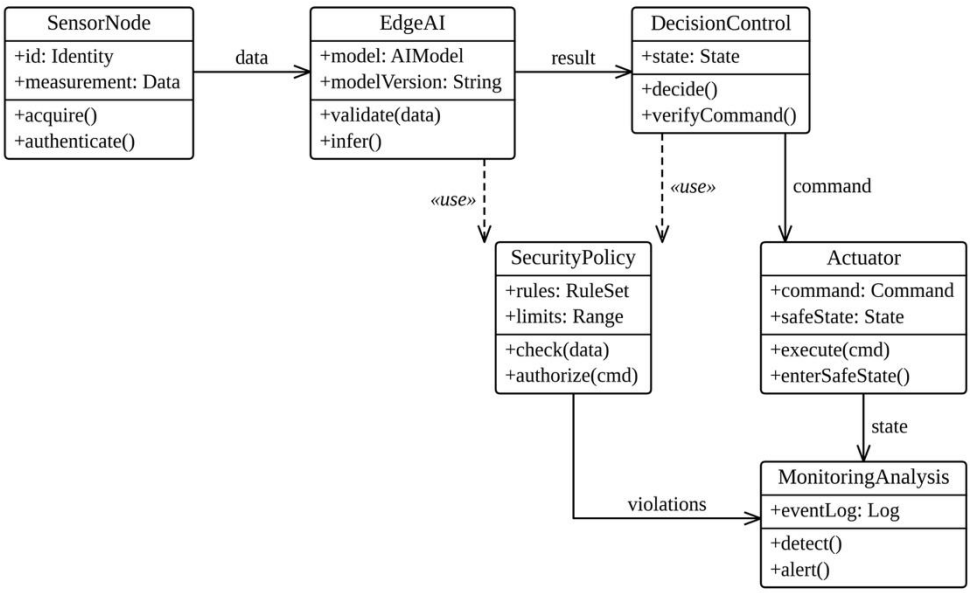


Fig. 3. UML class diagram of a protected autonomous system

The sequence diagram (Fig. 4) supplements the static model with a critical scenario. Incoming sensor data are checked against the permissible ranges defined in the security policy (check), after which EdgeAI generates a result (infer). Before execution, DecisionControl verifies the command against the SecurityPolicy (authorize). An authorized command is executed (execute), whereas in the event of denial, the system transitions to a safe state (enterSafeState) and MonitoringAnalysis generates an alarm (alert).

6. Risk Assessment Model

One of the key perspectives provided by the methodology under consideration is risk analysis. It serves as the basis for assessing assets, requirements, and protection measures.

For the purposes of this article, we examine a generalized quantitative model that allows for the comparison of threats without the need for complex probabilistic mechanisms [10].

For each threat T_i, a normalized probability P_i and impact I_i are assigned, and the baseline risk is:

$$R_i = P_i \cdot I_i \tag{1}$$

where $0 \le P_i, I_i, R_i \le 1$. The system risk is aggregated as a weighted sum:

$$R_{sys} = \sum_{i=1}^{n} w_i \cdot R_i, \quad \sum_{i=1}^{n} w_i = 1 \tag{2}$$

where w_i is a weight reflecting the importance of the respective asset or function. After applying a protection measure with effectiveness E_i, the residual risk is:

$$R_{res,i} = R_i \cdot (1 - E_i) \tag{3}$$

When m independent measures are applied to one threat, formula (3) generalizes to:

$$R_{res,i} = R_i \cdot \prod_{j=1}^{m} (1 - E_{ij}) \tag{4}$$

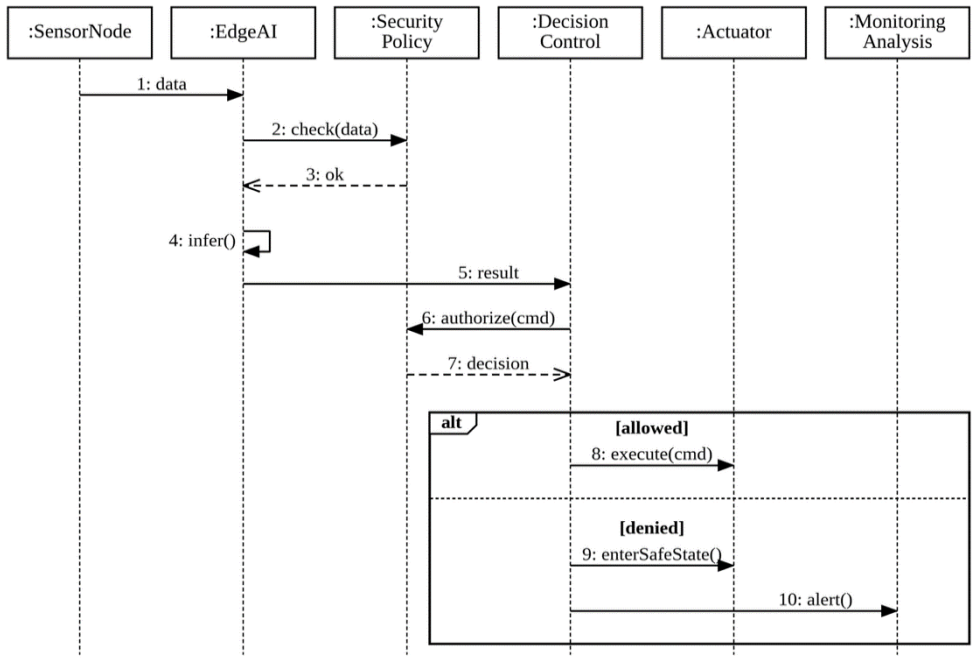


Fig. 4. UML sequence diagram for the execution of a critical command

The model is intended for prioritization rather than for obtaining absolute probabilities. The values are set by experts or derived from operational data. In Table 2, the effectiveness E of combined measures is given as a single expert estimate.

Table 2. Illustrative risk assessment for an autonomous mobile robot

Threat	P	I	R	Protection measure	E	R_{res}
Sensor data spoofing	0.60	0.90	0.54	Validation and authentication	0.65	0.19
Communication tampering	0.50	0.80	0.40	Encryption and segmentation	0.65	0.14
Unauthorized command	0.40	0.90	0.36	Access control and command checks	0.70	0.11
AI model manipulation	0.40	0.80	0.32	Version and integrity control	0.50	0.16
Malware on edge device	0.50	0.70	0.35	Endpoint protection	0.60	0.14
Denial of service (DoS)	0.60	0.60	0.36	Redundancy and monitoring	0.50	0.18
Unauthorized device	0.30	0.50	0.15	Device control	0.60	0.06
Configuration change	0.30	0.70	0.21	AAA and logging	0.60	0.08
System risk (equal weights)			**0.34**			**0.13**

With equal weights $w_i = 1/8$, the system risk is $R_{sys} = 0.34$ before and 0.13 after applying the measures, i.e. a reduction of about 61% (Fig. 5). The highest residual risk remains for sensor data spoofing and denial of service, which are candidates for additional measures.

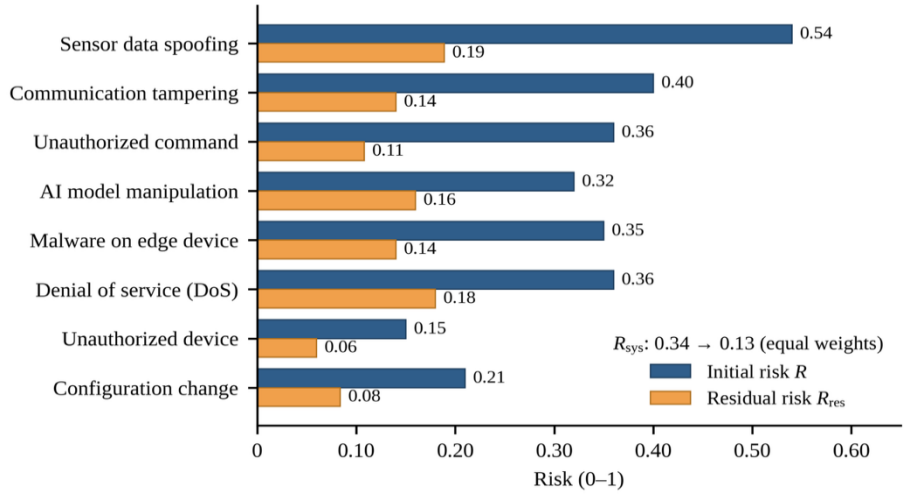


Fig. 5. Initial and residual risk by threat (according to Table 2)

7. Illustrative Industry 4.0 Scenario

To illustrate the application of the method in an Industry 4.0 environment, we consider the example of an autonomous mobile robot in a manufacturing setting. The robot utilizes various sensors—such as cameras—to perceive its surroundings, an edge computer with an AI module for decision-making, a controller and actuators, a wireless network for connectivity, and an operator interface (Fig. 6a). The configuration is typical of Industry 4.0, where IT and OT components are closely integrated, necessitating comprehensive protection for data, management systems, and physical processes.

As a first step in the management process, assets are defined based on sensor measurements, terrain recognition via an environment map, AI module models and parameters, movement commands, component status, configurations, production data, and logs. The boundaries between the field level, the edge level, and corporate or cloud services (B1 and B2 in Fig. 6a) are of particular importance. Each boundary represents a potential point for unauthorized access, data tampering, or the propagation of an attack into the system, corresponding to a channel between zones as defined in IEC 62443 [4].

The next step involves identifying vulnerabilities and formulating threat scenarios. Typical vulnerabilities include the lack of sensor source authentication, insufficiently secured communication, improper network segmentation, unprotected update interfaces, operator privilege escalation, and inadequate control over the AI model. Attack scenarios may involve sensor data spoofing, communication blocking, configuration changes, or unauthorized access to the management interface. The consequences of a successful attack propagate directly from the information environment to physical processes.

These risks necessitate the design and implementation of multi-layered protection comprising various measures to safeguard assets against specific threats. Examples include identification and access control, cryptographic communication protection, network segmentation, data integrity verification, and secure boot and software updates. Security measures must be considered holistically, as the compromise of one component can affect others. The values in Table 2 are illustrative and do not represent measured results from a specific robot.

Traceability between models is of particular importance (Fig. 6b). The "unauthorized command" threat from the risk model links to an access control requirement in the conceptual model, to the `SecurityPolicy` and `DecisionControl` classes in the UML model, to the `authorize` and `verifyCommand` functions in the implementation, and to a denied-command alarm in the monitoring system. Due to this interconnectedness, security becomes a fundamental property of the system architecture, rather than an additional function added after the system has been built.

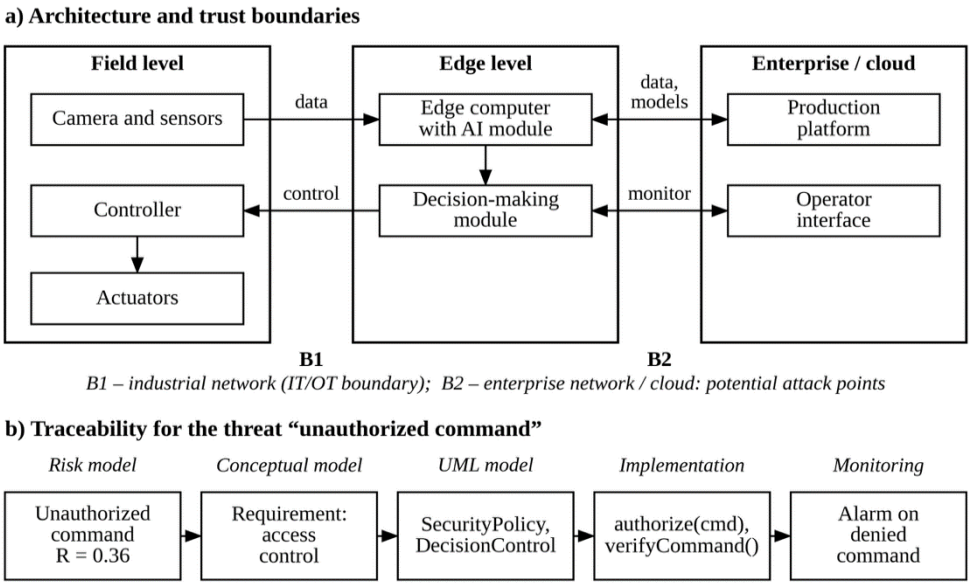


Fig. 6. Illustrative scenario: a) architecture of the autonomous robot and trust boundaries; b) traceability between the models

8. Summary of Security Requirements

The correspondence between viewpoints and the results is summarized in Table 3.

Table 3. Correspondence between viewpoints and modeling results

Viewpoint	Key question	Modeled object	Result
Information security	What can be breached?	Assets, threats, vulnerabilities	Threat model
Risk analysis	What is the priority?	P, I, R, residual risk	Protection priorities
Architecture	Where is protection implemented?	Components and relations	Conceptual model, class diagram
Functionality	How is it performed?	Interactions in critical scenarios	Sequence diagram
Technology	What is it implemented with?	Platforms, protocols, tools	Implementation model
Monitoring	How is it verified?	Events, alarms, indicators	Feedback

9. Discussion

The proposed method and corresponding models offer several practical advantages:

1. They are platform-independent. No specific platform needs to be selected during the design phase, allowing for various implementations of the same architecture depending on the specific production process scenario.

2. A universal communication language is established among various stakeholders: process engineers, legal experts, compliance specialists, and experts in security, automation, software, and maintenance.

3. The risk assessment model is directly linked to architectural and functional decisions. This is particularly important when mitigating attacks that originate in a communication or sensor component and culminate in a physical action within the production process.

A limitation of the current example is that the numerical estimates are based on expert judgment and serve illustrative purposes. Practical application requires real operational data, test results, logs, and incident information.

The next step involves transforming the UML model into an agent-based simulation model [11, 12], which would enable the study of threat dynamics and the impact of various security configurations.

Another area of focus is establishing automated traceability between requirements, risks, UML components, security measures, and monitoring results, ensuring that any architectural changes allow for the identification of affected requirements.

10. Conclusion

This paper presents a model-based approach to information security for automated and autonomous systems within Industry 4.0. The approach retains the core characteristics of the ISS development methodology: top-down design, conceptual modeling, and object-oriented transformation into a design model [2].

The approach enables the design of comprehensive protection for a production system composed of autonomous and automated systems.

The risk assessment model demonstrates how threats are prioritized and how protective measures relate to residual risk.

The paper examines an illustrative scenario in which system risk is reduced from 0.34 to 0.13.

This approach can serve as a foundation for more detailed models, simulations, and the implementation of security functions in Industry 4.0 automated and autonomous systems.

Acknowledgments

This work was supported by the NSP DS program, which has received funding from the Ministry of Education and Science of the Republic of Bulgaria under the grant agreement no. D01-74/19.05.2022

References

1. Thames, Lane & Schaefer, Dirk. (2017). Industry 4.0: An Overview of Key Benefits, Technologies, and Challenges. DOI: 10.1007/978-3-319-50660-9_1.
2. Gaidarski, I. K., Minchev, Z. B., Andreev, R. D.. Model Driven Approach for Designing of Information Security System. Journal of Information Assurance and Security, 2018, vol.13, MIR Labs, ISSN:1554-1010, 149-158

3. Schrecker S., H. Soroush, J. Molina et al., Industrial Internet of Things, Volume G4: Security Framework, Industrial Internet Consortium, 2016. (https://www.digitaltwinconsortium.org/pdf/IIC_PUB_G4_V1.00_PB.pdf)
4. IEC 62443-3-3:2013, Industrial communication networks – Network and system security – Part 3-3: System security requirements and security levels, IEC, 2013. (https://interoperable-europe.ec.europa.eu/collection/ict-standards-procurement/solution/iec-62443-3-32013-industrial-communication-networks-network-and-system-security-part-3-3-system)
5. Stouffer K., M. Pease, C. Tang, T. Zimmerman, V. Pillitteri, S. Lightman, A. Hahn, S. Saravia, A. Sherule, M. Thompson, Guide to Operational Technology (OT) Security, NIST Special Publication 800-82 Rev. 3, 2023 (https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-82r3.pdf)
6. IEC 61508:2010, Functional safety of electrical/electronic/programmable electronic safety-related systems, IEC, 2010. (https://interoperable-europe.ec.europa.eu/collection/ict-standards-procurement/solution/iec-61508-62010-functional-safety-electricalelectronicprogrammable-electronic-safety-related-systems)
7. ISO/IEC/IEEE 42010:2022, Software, systems and enterprise – Architecture description, ISO, 2022. (https://www.iso.org/standard/74393.html)
8. Object Management Group, OMG Unified Modeling Language (OMG UML), Version 2.5.1, OMG, 2017. (https://www.omg.org/uml/)
9. Dennis A., B. H. Wixom, D. Tegarden, Systems Analysis and Design: An Object-Oriented Approach with UML, 6th ed., John Wiley & Sons, 2020, ISBN: 978-1-119-56121-7, (https://www.wiley.com/en-us/systems-analysis-and-design-an-object-oriented-approach-with-uml-6th-edition-p-9781119561217)
10. National Institute of Standards and Technology, Guide for Conducting Risk Assessments, NIST Special Publication 800-30 Rev. 1, 2012. (https://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-30r1.pdf)
11. Gaidarski I., Kutinchev P., Transformation of UML Design Models of Information Security System into Agent-based Simulation Models, Information & Security: An International Journal 53, no. 1 (2022): 65-77 . https://doi.org/10.11610/isij.5305
12. Wooldridge M., An Introduction to MultiAgent Systems, 2nd ed., John Wiley & Sons, 2009. ISBN: 978-0-470-51946-2 (https://www.wiley.com/en-us/shop/general-introductory-computer-science/an-introduction-to-multiagent-systems-2nd-edition-p-9780470519462)

Insider Threats in Automated and Autonomous System Environments

Ivan Gaidarski

Laboratory of Unmanned Robotics Systems,
Institute of Robotics, Bulgarian Academy of Sciences, Sofia, Bulgaria
ivangaidarski@ir.bas.bg

Abstract. Autonomous and automated systems are becoming widespread across the economy - in logistics, manufacturing, agriculture, and energy. Traditional information security measures primarily target attacks from external sources and fail to account for environments where some insiders are machine identities, service accounts, and operators of control and update interfaces. This article extends an established insider threat model by adapting it for environments featuring automated and autonomous systems. It proposes a classification of threat actors, a generalized conceptual model for detection and protection, and a transformation of this model into an object-oriented UML design. A process for the automated detection of deviations from behavioral baselines is described, including the calculation of a composite risk score, with protection measures mapped to the stages of the attack lifecycle. The model's applicability is illustrated through examples from four sectors, and the method's limitations are discussed. The model is technology-agnostic and can serve as the foundation for a reference methodology for organizations implementing autonomous systems.

Keywords. Insider threats, insiders, autonomous systems, automated systems, machine identities.

1. Introduction

In recent years, autonomous and automated systems such as robots, industrial robotic lines, autonomous vehicles, unmanned aerial vehicles, and intelligent sensor networks have been widely adopted across the economy. Their widespread integration into the industry leads to increased productivity and reduced production costs. One of the most valuable advantages of these systems is their ability to perform operations in hard-to-reach and hazardous environments.

From the perspective of organizational information and cybersecurity, however, this leads to a significant expansion of the attack surface, increasing the likelihood of successful attacks and their associated consequences. Even a single compromised autonomous unit can cause physical damage, disrupt production processes, result in scrapped products and delayed deliveries, and lead to a loss of trust.

In information security, the classic approach is to protect against attacks from outside direction. Such are traditional approaches as perimeter protection, firewalls, intrusion detection systems (IDS) and intrusion prevention systems (IPS). Recently, special attention has been paid to internal threats posed by so-called "insiders"—employees or external service providers who have access to critical information resources and misuse them. Autonomous and automated systems pose a significant risk to information security, as they either have access to critical resources or constitute critical resources for the organization themselves. Typically, these systems are managed by internal personnel who could potentially become

malicious actors. This group includes teams responsible for maintenance, repair, and firmware updates, as well as operators and engineering teams; alongside them, software agents with machine identities are emerging, a trend that is rapidly accelerating with the adoption of artificial intelligence technologies. An additional complicating factor is that autonomous systems typically operate without direct human oversight, and deviations from their normal behavior are difficult to detect. These entities can be treated as insider actors because they operate within the trusted environment and may possess legitimate access to organizational resources.

This article proposes a model for detecting and defending against insider threats in environments featuring automated and autonomous systems. The model relies on a top-down approach, wherein a conceptual model of the domain is constructed and subsequently transformed into an object-oriented design model represented in UML. The conceptual model ensures technological independence, offering flexibility in the selection and implementation of specific security platforms.

The significance of the proposed model regarding the conference's thematic scope stems from its application in ICT-based economic activities, including logistics, manufacturing, agriculture, and energy. By linking machine identities, operational telemetry, and security policies, the model supports the secure operation and risk management of autonomous systems used in these sectors.

2. Related Work

Research into insider threats in information security examines the motivation, behavior, and consequences of actions taken by users with access to an organization's sensitive assets and resource - the so-called "insiders" [2, 11].

Risk analysis identifies intentional or unintentional malicious actions by insiders as a primary factor in security incidents, while simultaneously assessing the possibilities for mitigating the risk of data leakage or the compromising of assets resulting from successful attacks.

Automated detection of insider threats typically relies on profiling user actions and evaluating deviations from established baseline behavior [3, 12, 13]. Such studies primarily focus on traditional organizational environments where the actors are human. This article examines environments involving autonomous and automated systems, which present additional, specific risk factors. Examples include sensor data spoofing, unauthorized commands, compromise of the firmware updates chain, abuse of remote access, and attacks targeting decision making algorithms [6, 7]. Approaches based on the Zero Trust principle require every entity (human or machine) to be authenticated and authorized upon every access attempt within the environment [5].

Analyzing the attack lifecycle, combined with risk management models, provides a framework for structuring preventive and recovery measures [1, 4].

However, the reviewed approaches do not provide an integrated model that simultaneously considers human insiders, machine identities, autonomous entities, behavioral monitoring, and model-based security design. This article addresses this gap by offering the following contributions:

1. An extended classification of insider actors for automated and autonomous environments;
2. A technology-independent conceptual model for insider threat detection and protection;
3. A UML-based transformation of the conceptual model into a design model;
4. An automated detection and response process combining behavioral deviation, privilege level, and data sensitivity.

3. Foundations of Insider Threats in Automated and Autonomous Environments

The model is based on classic information security concepts - vulnerability, threat, source and agent of the threat, attack, data leak and incident.

A vulnerability is a weakness in the design, implementation, or operation of an asset that can be exploited by the source of the threat. A threat is the ability for a threat source to accidentally or intentionally exploit a specific vulnerability to cause harm. A threat agent is a person or process that transmits, causes, or sustains the threat. An attack is a targeted action that exploits a vulnerability to compromise the elements of the classic CIA triad - confidentiality, integrity or availability of an asset. A data breach is an incident where sensitive data becomes accessible to an unauthorized party [8].

In environments with automated systems, the concept of "insider" expands beyond the classical understanding. Along with people (employees, partners and suppliers), additional entities internal to the protected environment are added. These are system accounts used by technical and service teams, the machine identities of software agents communicating with each other, or update accounts. A compromised autonomous entity may act as an insider, like threat agent when it retains legitimate access to internal resources while operating under unauthorized control.

This necessitates adapting and expanding the classification of threat agents presented in Table 1.

Table 1. Categories of insider threats in environments with automated and autonomous systems

Actor	Type of access	Motivation	Example scenario
Privileged operator	Command interface of the autonomous unit	Revenge, financial gain, negligence	Changing a mission, exfiltration of mission data
Service account / technician	Maintenance, diagnostics, remote access	Malice, shared credentials	Installing unauthorized software, backdoor
Machine identity	M2M communication, API between agents	Compromise of a neighboring agent	Forged commands from a "trusted" system
Update provider	Software update channel	Economic espionage, bad faith	Injecting malicious code into an update
Compromised autonomous unit	Legitimate access to the internal network	External control (bot)	Using the unit as a foothold for an attack

The relationships between vulnerabilities, threats and attacks in automated environments are summarized in the model in Figure 1.

It preserves the structure of the classic model, adding some specific features to the classic elements of the vulnerability – threat – attack chain, such as telemetry and sensor data, software agents, machine identity and autonomous units.

4. Detection and Protection Model

The proposed model employs a top-down approach. Initially, an analysis of the problem domain is conducted to identify all stakeholders. The next step involves constructing a conceptual model based on this analysis, which is subsequently transformed into an object-oriented design model. This sequence ensures technological independence, allowing the same

model to be supported by various protection mechanisms [9]. The model addresses four fundamental questions: what, why, how and where is being protected.

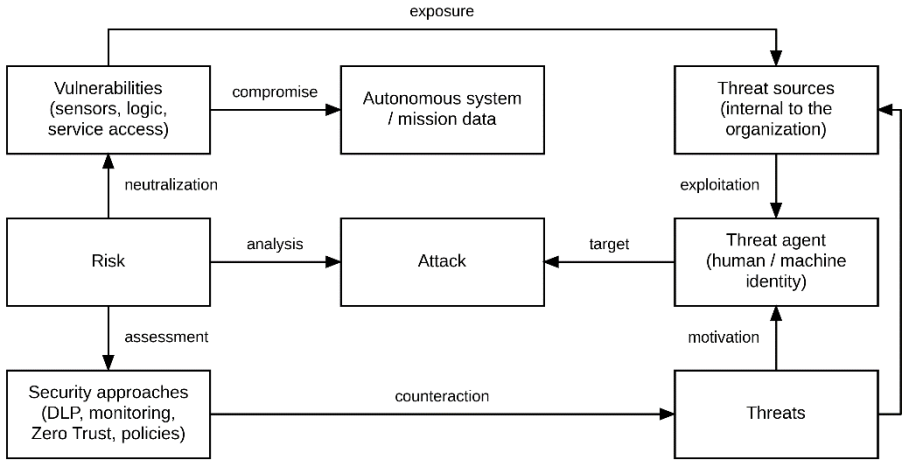


Fig. 1. Model of the vulnerability – threat – attack chain in environments with automated and autonomous systems

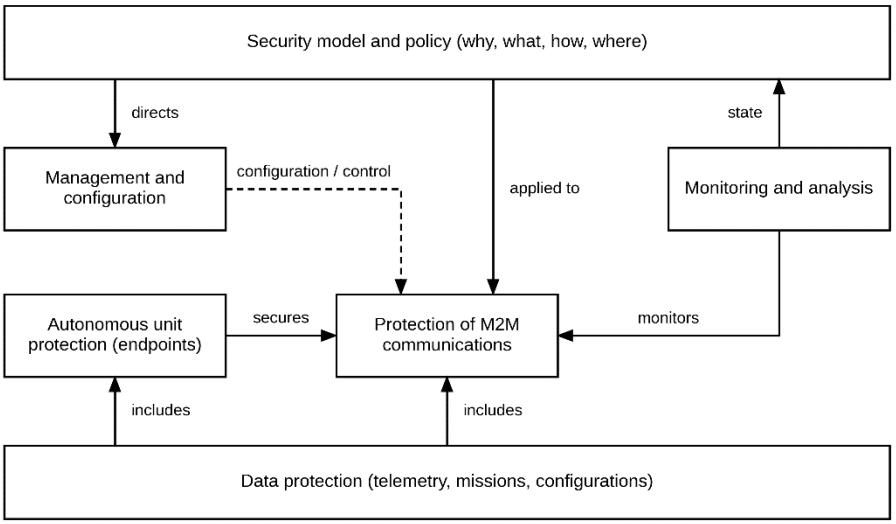


Fig. 2. Generalized conceptual model for detection and protection against insider threats in environments with autonomous systems

The generalized conceptual model (Figure 2) consists of six components. The "Security Model and Policy" component reflects the environment, including regulatory and technological requirements, industry standards, and internal organizational policies. It defines which operations are critical and/or sensitive, as well as who has access to them and under what conditions. The "Management and Configuration" component assigns tasks to the protection components and enforces the applicable security policies. The "Monitoring and Analysis" component tracks the settings of the other components, collects telemetry and logs,

detects deviations from the baseline. The components "Protection of Autonomous Units," "Protection of M2M Communications," and "Data Protection" provide security capabilities for endpoints, machine-to-machine (M2M) traffic, and data, respectively covering the three primary data states: data at rest, data in motion, and data in use [8].

Using the object-oriented description language UML, the conceptual model is transformed into an object-oriented design model. The different types of diagrams from the UML arsenal allow the description of architecture, structure, processes, workflows, interaction with other objects and systems, as well as representation of various processes through time diagrams.

Figure 3 presents a UML "Class-Diagram" describing the structure of the design model, preserving the logic of the conceptual model. Four main classes are defined for this purpose. The "Management" class represents the security model and policy and manages the configuration. The "Protection" class performs the protection of autonomous units and communications. The "Control" class provides telemetry monitoring, logs, baseline deviations, and access control on the various interfaces. The "Data" class implements the solutions to skip, block, isolate and report sensitive content [10]. The design model thus defined is independent of the specific platform and can be a guide in the selection of existing data leakage prevention systems, for example Data Leak Prevention (DLP) solutions.

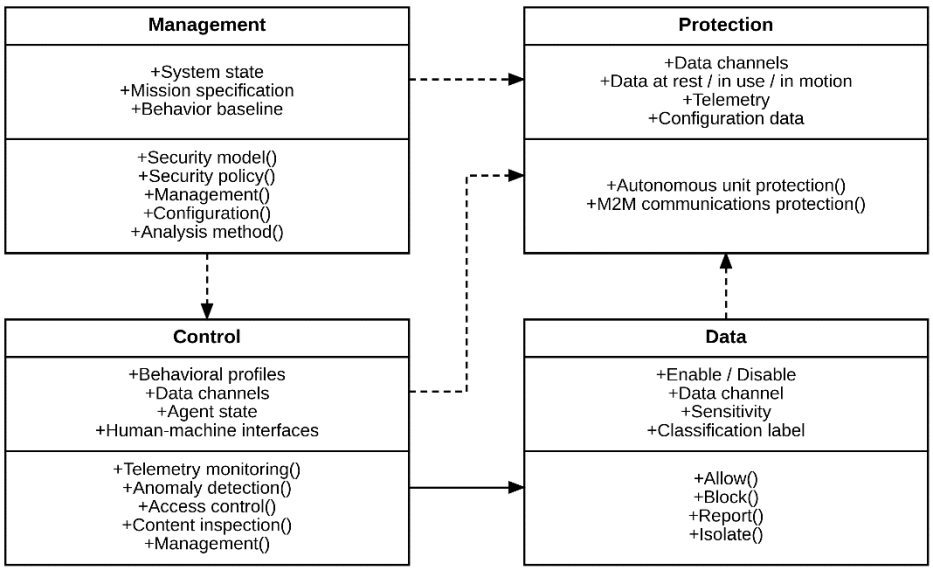


Fig. 3. UML class diagram of the detection and protection model

5. Automated Detection Process

Due to the substantial volume of telemetry data and logs, the process of detecting insider threats in automated environments requires the implementation of automation.

A model for the automated detection, assessment, and response to insider threats is presented in Figure 4. The process begins with the collection of telemetry data from autonomous units, command interfaces, and access points. Following filtering and normalization, the observed metrics are compared against a baseline of normal behavior established during training in a trusted environment.

The deviation of an indicator from the baseline is expressed in a generalized way by the standardized difference:

$$D(x) = |x - \mu_b| / \sigma_b \quad (1)$$

where μ_b and σ_b are, respectively, the mean value and the standard deviation of the baseline. The risk of a given event is assessed in a generalized manner as the product of the probability of its occurrence and its impact:

$$R = P(E) \cdot I \quad (2)$$

For each agent (user, account, or autonomous entity), an insider threat risk score is calculated, combining behavioral deviation, agent privileges, and the sensitivity of the accessed data:

$$IRS(a) = w_1 \cdot D(a) + w_2 \cdot P(a) + w_3 \cdot S(a), \quad w_1 + w_2 + w_3 = 1 \quad (3)$$

where $D(a)$ is the deviation, $P(a)$ is the relative privilege, $S(a)$ is the sensitivity of the data, and the weights w_1, w_2, w_3 are determined by the security policy of the organization. Based on thresholds τ_1 and τ_2, an action is selected:

$$\begin{aligned} Action(a) = \text{block, if } IRS \geq \tau_2; \\ \text{isolate, if } \tau_1 \leq IRS < \tau_2; \\ \text{monitor, if } IRS < \tau_1 \end{aligned} \quad (4)$$

The key indicators used for detection are summarized in Table 2. Following each event, the policy and the baseline are updated, thereby closing the management cycle.

Table 2. Indicators for automated detection of insider threats

Indicator	Source	Normal behavior	Alarming deviation
Mission route and time	Telemetry from the autonomous unit	Preset route	Deviation without authorization
Volume of outgoing data	Network logs	Within the profile limits	Sudden increase in traffic to external addresses
Frequency of service access	Account logs	Planned interventions	Access outside working hours
Commands from M2M interfaces	API logs	Expected requests between agents	Commands from an unknown agent
Software change	Update register	Signed and approved updates	Unsigned or unexpected code

6. Mapping to the Attack Life Cycle

The effectiveness of defensive measures can be verified by matching them to the stages of the attack life cycle [1]. Figure 5 shows the main stages of the attack – reconnaissance, arming, delivery, exploitation, installation, takeover (command and control) and defensive actions against the attack targets, as well as the main countermeasures in the context of autonomous and automated environments.

Key countermeasures include the principle of least privileges for all accounts, Zero Trust identities, channel filtering, endpoint protection and segmentation, software integrity control at boot, signing updates, and applying DLP to mission data [5]. Neutralizing an attack earlier reduces the impact and cost of response.

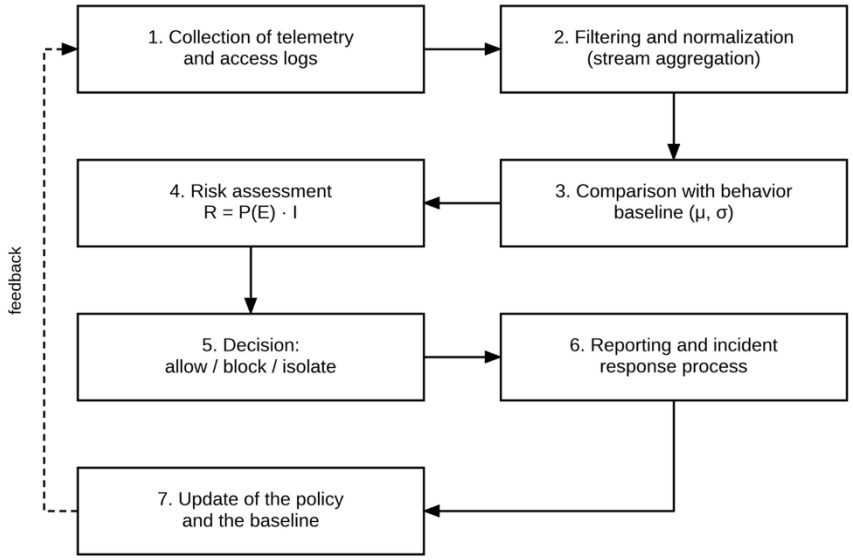


Fig. 4. Process of automated detection, assessment and response to insider threats

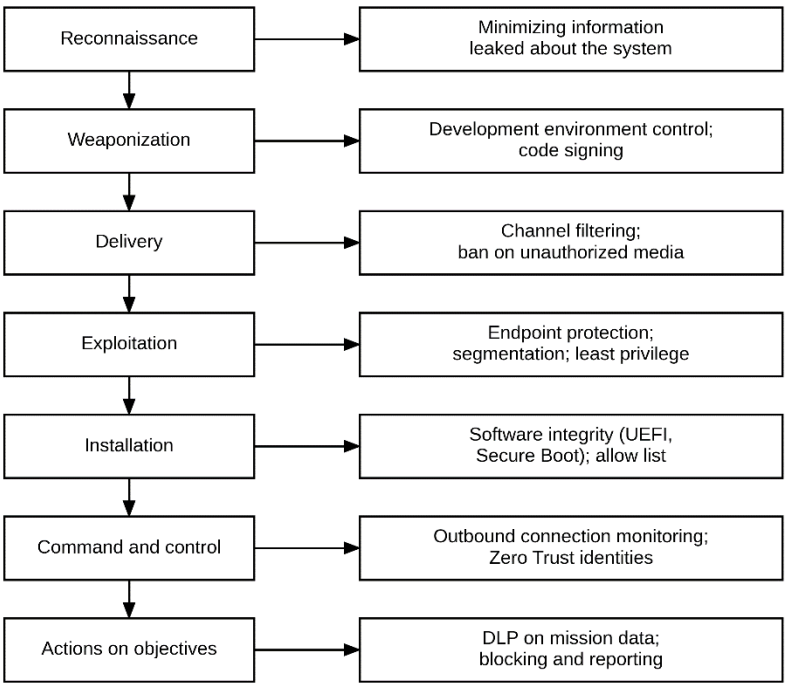


Fig. 5. Stages of the attack life cycle and main countermeasures in environments with autonomous systems

7. Applicability and Limitations

Applicability

The proposed model is applicable in any sector implementing autonomous or automated systems, where part of the process involves the execution of critical operations by insiders - operators, service accounts, machine identities and update providers. Since the model is technology independent, the organization can choose specific platforms for implementation, according to the technology level, scope and cost of implementation. Such platforms can be a set of traditional security measures, software agents to protect endpoints, DLP solutions to prevent data leakage, without having to change the design model. In practice, industry-specific templates can be used.

The proposed model may support the implementation of selected technical and organizational controls relevant to frameworks such as the EU GDPR data protection regulation [14], network and information systems regulation NIS2 [15], The Cyber Resilience Act (CRA) [16], Sarbanes-Oxley Act (SOX) [17], Health Insurance Portability and Accountability Act (HIPAA) [18] and Payment Card Industry (PCI) Data Security Standard (DSS) [19].

Table 3 shows applications of the paper's model in four sectors, matching each scenario with an actor from Table 1, indicators from Table 2, and countermeasures from Figure 5. The scenarios themselves are hypothetical and illustrative. They do not represent the result of implementation in a real organization.

Table 3. Illustrative scenarios for applying the model by sector

Sector and system	Actor (Table 1)	Indicators (Table 2)	Main countermeasures
Logistics and warehouses: fleet of autonomous mobile robots	Service account / technician	Frequency of service access; software change	Least privilege for service accounts; signed updates
Agriculture: unmanned aerial vehicles	Privileged operator	Mission route and time; volume of outgoing data	DLP on mission data; monitoring of outbound connections
Energy: automated equipment	Machine identity; update provider	Commands from M2M interfaces; software change	Zero Trust identities; segmentation; software integrity at boot
Manufacturing: robotic lines	Compromised autonomous unit	Commands from M2M interfaces; volume of outgoing data	Segmentation; isolation of the unit at $IRS \geq \tau_1$

In the field of education, the model can serve as a basis for training cybersecurity specialists. The conceptual model and object-oriented UML transformations are part of model-based design - a skill essential for creating information security systems.

The attack lifecycle model and fundamental defensive measures are integral components of information security training. The model for the automated detection, assessment, and response to insider threats is suitable for practical exercises in simulation environments, where students can generate test data and define response thresholds [10].

Limitations

The proposed method is conceptual and its limitations should be kept in mind when interpreting the findings. They relate to the formulas, to the baseline, to the life cycle of the attack and to the practical implementation.

The article does not contain experimental data and the conclusions about applicability and compliance with the attack life cycle are logical, not empirically proven. The model has not been validated with a real data set, no universal values for the weights and thresholds have been defined. Detection accuracy and false positive rate were not measured.

Therefore, the model should be considered as a basis for design. The next step is experimental verification through a simulation or laboratory environment. Normal and abnormal scenarios need to be generated and accuracy, sensitivity and response time evaluated

The baseline is built in a trusted environment. If an insider has already acted maliciously during the training period, his behavior is perceived as normal. The classical life cycle stages are described for external attackers. Domestic actors already have legitimate access, so intelligence, weaponization, supply, and exploitation can be eliminated or severely curtailed, and countermeasures at these stages (Figure 5) do not cover the abuse of rights already granted. The comparison in Section 6 is qualitative and does not prove effectiveness.

The classification of actors in Table 1 is not exhaustive and does not cover all possibilities, for example collusion between insiders or autonomous AI agents whose behavior is more difficult to profile. The claim that the model can serve as the basis of a reference methodology remains a hypothesis pending its experimental validation.

8. Conclusion

The article extends the insider threat model to environments featuring automated and autonomous systems. In addition to humans, insider actors include service accounts, machine identities, and compromised autonomous entities. The key results are:

- Classification of threat actors, including machine identities and compromised autonomous entities;
- Generalized conceptual model for detection and protection, transformed into an object-oriented UML design model;
- Automated detection process based on a behavioral baseline and a generalized risk assessment;
- Mapping of measures to the attack lifecycle.

The model is conceptual and has not been experimentally validated.

Limitations regarding the formulas, the baseline, and the attack lifecycle mapping are discussed in Section 7.

The proposed model links security policy, monitoring, risk assessment, and defensive actions. It can serve as a basis for designing defenses, though experimental validation is required for practical application. The model is technology-agnostic and can provide a conceptual basis for developing security methodologies for organizations deploying autonomous and automated systems.

Future work involves experimental validation through multi-agent simulation, including test data generation and the calibration of weights and thresholds, the application of machine learning methods for anomaly detection, and the extension of the model to include software supply chain security components.

Acknowledgments

This work was supported by the NSP DS program, which has received funding from the Ministry of Education and Science of the Republic of Bulgaria under the grant agreement no. D01-74/19.05.2022.

References

1. Hutchins E., M. Cloppert, R. Amin, Intelligence-Driven Computer Network Defense Informed by Analysis of Adversary Campaigns and Intrusion Kill Chains. Proc. of the 6th International Conference on Information Warfare and Security, 2011, pp. 113–127.
2. Cebula J., M. Popeck, L. Young, A Taxonomy of Operational Cyber Security Risks Version 2, Technical Report CMU/SEI-2014-TN-006, Software Engineering Institute, Carnegie Mellon University, 2014, https://doi.org/10.1184/R1/6571784.v1.
3. Kandias M., A. Mylonas, N. Virvilis, M. Theoharidou, D. Gritzalis, An Insider Threat Prediction Model. In: Katsikas S., J. Lopez, M. Soriano (eds) Trust, Privacy and Security in Digital Business (TrustBus 2010), Lecture Notes in Computer Science, vol. 6264, Springer, Berlin, Heidelberg, 2010, pp. 26–37, https://doi.org/10.1007/978-3-642-15152-1_3.
4. ISO/IEC 27005:2022, Information security, cybersecurity and privacy protection — Guidance on managing information security risks, International Organization for Standardization, 2022, https://www.iso.org/standard/80585.html.
5. Rose S., O. Borchert, S. Mitchell, S. Connelly, Zero Trust Architecture, NIST Special Publication 800-207, National Institute of Standards and Technology, 2020, https://doi.org/10.6028/NIST.SP.800-207.
6. ENISA, ENISA Threat Landscape 2023, European Union Agency for Cybersecurity, 2024, https://www.enisa.europa.eu/publications/enisa-threat-landscape-2023.
7. MITRE, ATT&CK® Knowledge Base, The MITRE Corporation, 2024, https://attack.mitre.org.
8. Gaidarski I., Z. Minchev, Insider Threats to IT Security of Critical Infrastructures. In: Tagarev T., K. T. Atanassov, V. Kharchenko, J. Kacprzyk (eds) Digital Transformation, Cyber Security and Resilience of Modern Societies, Studies in Big Data, vol. 84, Springer, Cham, 2021, pp. 381–394.
9. Gaidarski I., A. Madzharov, Applying a New Approach to Consider the Human Factor in the Design of Information Security Systems. Information & Security: An International Journal, vol. 55, no. 3, 2024, pp. 261–272.
10. Gaidarski I., P. Kutinchev, An Approach for Constructing a Simulation Model for Dynamic Analysis of Information Security System. Lecture Notes in Networks and Systems, vol. 418, Springer, Cham, 2022, pp. 518–526.
11. Renaud K., M. Warkentin, G. Pogrebna, K. van der Schyff, VISTA: An Inclusive Insider Threat Taxonomy, with Mitigation Strategies. Information & Management, vol. 61, no. 1, 2024, Art. 103877, https://doi.org/10.1016/j.im.2023.103877.
12. Pfleeger S. L., D. D. Caputo, Leveraging Behavioral Science to Mitigate Cyber Security Risk. Computers & Security, vol. 31, no. 4, 2012, pp. 597–611, https://doi.org/10.1016/j.cose.2011.12.010.
13. Legg P. A., N. Moffat, S. Happa, I. Agrafiotis, M. Goldsmith, S. Creese, Automated Insider Threat Detection System Using User and Role-Based Profile Assessment. IEEE Systems Journal, vol. 11, no. 2, 2017, pp. 503–512, https://doi.org/10.1109/JSYST.2015.2438442.
14. European Parliament and Council, Regulation (EU) 2016/679 (General Data Protection Regulation - GDPR), Official Journal of the European Union, L 119, 2016, https://eur-lex.europa.eu/eli/reg/2016/679/oj.
15. European Parliament and Council, Directive (EU) 2022/2555 (NIS2 Directive on Measures for a High Common Level of Cybersecurity Across the Union), 2022, https://digital-strategy.ec.europa.eu/en/policies/nis2-directive.
16. European Commission, Regulation on Horizontal Cybersecurity Requirements for Products with Digital Elements (Cyber Resilience Act), 2024, https://digital-strategy.ec.europa.eu/en/policies/cyber-resilience-act.
17. Belina H., D. V. Rama, Early Warnings of SOX 404 Material Weaknesses in Internal Control. Current Issues in Auditing, vol. 16, no. 2, 2022, pp. A1–A9, https://doi.org/10.2308/ciia-2021-030.
18. Weiss J. N., The Health Insurance Portability and Accountability Act (HIPAA). In: Physician Crisis, Springer, Cham, 2023, pp. 215–228, https://doi.org/10.1007/978-3-031-27979-9_15.
19. Lincke S., Complying with the PCI DSS Standard. In: Information Security Planning, Springer, Cham, 2024, pp. 45–62, https://doi.org/10.1007/978-3-031-43118-0_3.

Opportunities for Using Digital Twins in Natural Disaster Risk Management

Hristina Angelova

University of National and World Economy, Sofia, Bulgaria
hristina.lesova@unwe.bg

Abstract. The rapid transition from Industry 4.0 to Industry 5.0 has changed how we manage complex smart systems, making real-time data processing more critical than ever. This paper investigates the conceptual nature, architectural characteristics, and practical applications of Digital Twin technology, with a specific focus on innovative disaster risk management in modern urban ecosystems. The study establishes a precise academic distinction between digital twins and conventional computer simulations based on the criterion of an automated, bidirectional data flow. Furthermore, the main technological challenges hindering the deployment of these systems are analyzed, including Big Data management, lack of interoperability standards, and cybersecurity vulnerabilities. Through an analysis of leading global practices from Singapore and the Netherlands, the paper demonstrates how city-scale digital twins transform civil protection from a reactive to a proactive workflow. The article outlines a theoretical and applied framework for building cognitive urban resilience during crises, proving that digital twins are a critical tool for preventing and mitigating the compounding damages of climate-induced disasters.

Keywords. Digital twin, Industry 5.0, Natural disaster, Risk management, Smart Cities, Urban Resilience.

1. Introduction

The current stage of development in the global scientific and industrial sphere is marked by a profound digital transformation. The introduction of the Industry 4.0 concept and the rapid transition to Industry 5.0 are changing the way complex systems are managed and optimized. At the core of this evolution lies the massive deployment of so-called cyber-physical systems (CPS), which integrate physical processes with computational algorithms and network structures. As these systems become increasingly complex, the scientific literature highlights the need for a closer and more dynamic connection between the real and virtual worlds: *"With the advent of Industry 4.0, cyber-physical systems (CPS) have become the core of smart manufacturing. However, the increasing complexity of these systems demands a more integrated approach to connect the physical and virtual spaces in real time."* [1].

The growing complexity of cyber-physical environments presents researchers with a serious challenge: how to achieve continuous optimization, analysis, and management without requiring the interruption of real, often socially critical processes. In this highly dynamic environment, traditional approaches relying on historical data or static simulations no longer possess the necessary flexibility. Real-time data processing is becoming a critical factor for security and efficiency. *"A primary driver for Digital Twin technology is the requirement to optimize complex operations without disrupting ongoing physical processes. Real-time data feeds enable continuous calibration of the virtual model to mirror the exact state of its physical counterpart."* [2].

The transition to Industry 5.0 further reinforces this necessity by expanding the scope of technologies beyond pure machine automation and focusing on the human factor, sustainable development, and the ability of systems to survive and recover from major shocks. According to official reports: *"Industry 5.0 complements the existing Industry 4.0 paradigm by highlighting research and innovation as drivers for a transition to a sustainable, human-centric and resilient industry, moving the focus from shareholder to stakeholder value."* [3]. This requires the creation of intelligent tools that help people make fast and correct decisions in emergency situations.

Therefore, the relevance of the problem under consideration is rooted in the urgent need to transition from previous methods to proactive and predictive risk management. The digital twin is establishing itself as a natural and necessary response to the challenges of Industry 4.0 and 5.0, as it allows the simulation of dangerous scenarios and natural disasters in a safe virtual environment, ensuring the protection of infrastructure and human life.

The topicality of the problem is reinforced by alarming global statistics, which report an unprecedented growth in the frequency and intensity of natural disasters in recent years. According to current reports by the World Meteorological Organization (WMO), the past decade is the hottest in modern measurement history, which directly reflects on the acceleration of extreme weather phenomena. In their global analyses, scientists and climatologists confirm the scale of the crisis: *"On a day-to-day basis, our weather has become more extreme. Heatwaves, wildfires, drought, tropical cyclones, storms and flooding are causing thousands of deaths, impacting millions of people and causing billions in economic losses, highlighting the vulnerability of our inter-connected economies."* [4]

This informational and methodological shortcoming is deepened by the fact that modern cities function as hyper-connected macrosystems. When a natural disaster, such as a flash flood, affects a given urban area, it does not cause an isolated problem, but generates a cascading chain of failures that simultaneously impact social, technical, and economic structures. Traditional geographic information systems offer a static view of topography, but remain blind to dynamic changes in traffic behavior, local power outages, or fluctuations in the hydraulic pressure of protective facilities at that exact moment.

This growth is also confirmed by statistical data from the International Disaster Database (EM-DAT). These data show that hydrometeorological events - specifically flash floods and extreme storms—now account for over 70–80% of all reported natural disasters on a global scale. Risk management analysts point out that the financial toll of these crises is increasing significantly: *"Compared to the long-term historical averages, the frequency of global natural disasters, particularly floods and extreme storm events, has shown an upward trend, driving direct economic losses up by over 30% to 40% in recent consecutive years."*. [5] It is precisely this drastic increase in risks that requires the abandonment of previous static response models and the urgent transition to digital twins for urban planning and civil protection.

The resolution of this deep research problem is scientifically possible only through the implementation of digital twins, which provide the sought-after technological foundation for innovative and proactive risk management. By creating a living digital counterpart of the urban environment, continuously calibrated by sensor data, the technology enables the closure of the "observation - simulation – decision - action" loop. This makes it possible, for the first time in the history of civil protection, to run thousands of parallel disaster simulations in real time, discovering the optimal path for evacuation and the distribution of rescue resources even before the critical point has occurred in the physical world. At the core of this paper lies the key research problem associated with the acute need to transition to adequate, intelligent, and innovative management of natural disaster risks in urban ecosystems. Traditional civil

protection approaches, which rely primarily on reactive measures after an event occurs or on linear evacuation plans, no longer correspond to the dynamics of modern climate threats. Security researchers emphasize that the lack of integration between simulation models and real-time world data hinders effective decision-making:*"The core research gap in current disaster risk management lies in the fragmented nature of data and models, which limits proactive risk assessment and real-time operational response during extreme weather events."*. [6] Therefore, the scientific and practical challenge consists in developing dynamic frameworks that can predict not only the disaster itself, but also its cascading effects on critical infrastructure (roads, power grids, hospitals). Innovative management requires systems capable of learning and adapting autonomously. As modern studies dedicated to emergency resilience note: *"Innovative risk management demands shifting from static hazard mapping toward dynamic, intelligence-driven systems that can simulate compounding disaster scenarios and automatically optimize mitigation strategies to protect human lives.".[7]* The resolution of this research problem is possible only through the implementation of digital twins, which provide the sought-after technological foundation for innovative and proactive risk management.

The main objective of this publication is to explore and systematize the possibilities for integrating digital twin technology into disaster risk management processes in urbanized areas. In this way, a conceptual framework is outlined for the transition from traditional methods to innovative, proactive, and intelligent management of urban resilience during crises. The object of the study is digital twins as high-tech socio-technical systems for real-time analysis, while the subject of the study covers their specific functional capabilities for forecasting, scenario simulation, and the optimization of evacuation and rescue plans in emergency situations.

Thus, the study aims to outline a conceptual theoretical-applied framework for the transition from traditional, reactive methods to innovative, proactive, and intelligent management of urban crisis resilience. The expected scientific and practical contribution of the article is expressed in the systematization of the technological requirements for building a sustainable civil protection architecture, which can serve as a basis for future developments in the field of autonomous management of urban regions at risk.

2. Nature and definition of the digital twin concept

To understand the scientific and practical significance of digital twins, it is necessary for the concept to be precisely defined and differentiated from traditional computer modeling methods. In the most general sense, a digital twin represents a virtual engineering of a real object, process, or system that does not merely depict it visually, but replicates its behavior and state in real time through a two-way data flow. Scientific literature points out that this technology consists of a physical layer, a virtual layer, and connecting algorithms that unite them: *"The Digital Twin is defined as a virtual representation of a physical object or system that scenarios its lifecycle, using real-time data to enable understanding, learning and reasoning."* [8].

One of the most common mistakes in academic papers is confusing a digital twin with ordinary 3D simulations or CAD (Computer-Aided Design) models. The main difference is rooted in the nature of data flow automation. While traditional simulation operates in a closed virtual environment with predefined static parameters, a digital twin requires continuous and bidirectional integration with the real world. Leading researchers in the field define the levels of digitization: *"We distinguish between a Digital Model, a Digital Shadow, and a Digital Twin based on data integration. Only a Digital Twin features a fully automated, bidirectional data flow where changes in the physical object directly impact the virtual state, and vice versa."* [9].

This two-way integration means that sensors located on the physical object continuously send real data to the virtual model, which in turn processes this data through algorithms and sends commands back to the physical components to optimize operation or prevent accidents. From a lifecycle perspective, the digital twin is not a static software product, but a dynamically evolving tool. It accompanies its physical counterpart from the design stage, through operation, to its decommissioning. The scientific literature emphasizes that this dynamic nature requires a stable computing architecture: *"A Digital Twin is an integrated multiphysics, multiscale, probabilistic simulation of an asset or system that uses the best available physical models, sensor updates, and history to mirror the life of its corresponding twin."* [10]. It is precisely this ability to capture the entire lifecycle in real time that makes the technology applicable in the management of complex and unpredictable macrosystems, such as urban environments and critical infrastructure during emergency situations.

3. Areas of application and features of digital twins

The practical implementation of digital twins covers a wide spectrum of high-tech sectors, becoming a primary catalyst for innovation and the transition to intelligent resource management. The current decade marks a transition from conceptual developments to mass deployment of the technology in high-criticality industries. Global studies show that the scope of application is transformative: *"The application of Digital Twins has expanded far beyond aerospace and manufacturing into domains such as healthcare, smart cities, energy management, and automotive industries, transforming them into data-driven ecosystems."* [2].

Despite the enormous advantages, the implementation of digital twins is associated with specific technological and structural features that pose serious challenges to modern science. Since these systems require a continuous real-time data transfer from thousands of IoT devices, the first main feature is the need for management and lightning-fast analysis of large data sets (Big Data), combined with the need for enormous computing power. The second critical challenge is the lack of standardization and low interoperability between different software and hardware platforms.

Thirdly is the issue of data security and cybersecurity of the virtual environment. Due to the nature of the two-way data flow, any compromised point in the digital model can lead to real physical damage. Researchers pay exceptionally serious attention to these risks: *"The integration of Digital Twins opens significant challenges regarding data security and privacy. Since a Digital Twin holds a complete, real-time digital replication of a physical asset or critical infrastructure, any cyber-attack targeting the virtual space can have catastrophic physical consequences."* [11]. In view of these features, the successful implementation of digital twins requires the deployment of modern encryption protocols, edge computing, and robust cybersecurity architectures, especially when the technology is applied to the management of critical urban infrastructure.

4. Application of digital twins in natural disaster management

The most critical and socially significant direction in the development of digital twins is their integration into the management of urban environments, the so-called Smart Cities, and territories for the purpose of prevention, simulation, and response to emergency situations and natural disasters. In the conditions of intensifying climate change, traditional crisis management plans based on static maps prove to be ineffective. Digital twins of cities allow operators to monitor geographical, meteorological, and infrastructural data in real-time, predicting the development of a disaster seconds or hours before it affects the population. Scientific research in this field shows that: *"Digital twins can combine diverse datasets to*

provide real-time situational awareness and predictive simulations for disaster response, allowing decision-makers to test evacuation strategies and resource deployment in a risk-free virtual environment before real-world execution." [12].

The practical application of this technology can be observed in advanced digital nations that use large-scale three-dimensional digital replicas of their territories. A global example is Singapore with its Virtual Singapore project. The city-state uses its digital twin to simulate various natural and anthropogenic crises—ranging from the spread of fires in densely populated areas to predicting evacuation routes during earthquakes or extreme winds. A similar example is the Netherlands, specifically the city of Rotterdam, where the digital twin is a vital tool for the survival of low-lying zones. By integrating data from hydrological sensors, groundwater level sensors, and sea-level sensors, the system simulates potential floods during dike breaches or torrential rains, automatically activating protective barriers and informing rescue teams.

In the context of national development, the application of urban digital twins for crisis management and natural disaster prevention in our country is predominantly at the stage of local pilot projects and academic initiatives, while a large-scale national coordination framework is still lacking. An example of the application of digital twins in Bulgaria is the GATE pilot project "Digital Twin of the City," which is a major interdisciplinary pilot project developing on the territory of the "Lozenец" District in Sofia with the support of Sofia Municipality. "The creation of the digital twin includes the construction of a 3D model that represents the semantics and geometry of buildings, green areas, road infrastructure, and other urban environment objects. The visualization of the 3D model is an example of a multi-dimensional representation of the data available for the city in an interactive and easily digestible way for stakeholders. The 3D model is continuously enriched with new data, allowing the implementation of various scenarios related to urban processes and phenomena" [14]. The technology aggregates and analyzes urban data to simulate and visualize dynamic processes in a real environment. Its main goal is to provide an intuitive tool for testing, planning, and optimization that assists Sofia Municipality in making managerial decisions based on empirical data. This shows that conceptual models for integrating geospatial data are being created at the level of higher technical schools and scientific organizations, although their practical implementation in operational civil protection headquarters remains limited.

5. Conclusion

The implementation of digital twin technology represents a key stage in the evolution of modern digital transformation, marking the transition from passive automation to intelligent and proactive lifecycle management of complex cyber-physical systems. Within the framework of this research, the fundamental essence of the digital twin was examined, making a clear scientific distinction between the concept and traditional static computer simulations based on the criterion of an automated, two-way real-time data flow.

The analysis of architectural features highlights that although the technology's potential is transformative for industries such as manufacturing, healthcare, and energy, its successful scaling requires overcoming serious challenges related to interoperability, big data processing, and guaranteeing the cybersecurity of critical infrastructure.

Particularly significant is the role of digital twins in the context of Industry 5.0, where the focus shifts toward societal resilience during crises. Practical examples from cities like Singapore and Rotterdam categorically prove that city-scale digital twins completely change the philosophy of emergency management and natural disaster response. Through the integration of IoT sensors and artificial intelligence-based predictive analytics, risk

management shifts from a reactive mode of action to a proactive phase of simulation, prevention, and real-time damage minimization.

The digital twins are no longer merely a conceptual engineering tool, but a strategic necessity for building secure, resilient, and flexible socio-technical ecosystems capable of responding to the dynamic challenges of the future and climate change. Future scientific research in this field should focus on the standardization of data protocols and the development of fully autonomous cognitive digital twins for civil protection needs.

Acknowledgments

This work was financially supported by the UNWE Research Programme under the project "Application of Artificial Intelligence for Sustainable Business Development in the Context of Climate Change" (Research Grant No. NID NI -35/2025/A).

References

1. Tao, F., Zhang, H., Liu, A., Nee, A. Y. (2018). Digital twin in industry: State-of-the-art. IEEE Transactions on Industrial Informatics. https://www.researchgate.net/publication/320875040_Taking_Liberties_with_the_Past_Monuments_Memorials_and_Memory
2. Fuller, A., Fan, Z., Day, C., Barlow, C. (2020). Digital Twin: Enabling Technologies, Challenges and Open Research Directions. IEEE Access. https://ieeexplore.ieee.org/document/9103025
3. Breque, M., De Nul, L., Petridis, A. (2021). Industry 5.0: Towards a sustainable, human-centric and resilient European industry. European Commission, Directorate-General for Research and Innovation, Publications Office of the European Union. https://research-and-innovation.ec.europa.eu/knowledge-publications-tools-and-data/publications/all-publications/industry-50-towards-sustainable-human-centric-and-resilient-european-industry_en
4. World Meteorological Organization (2026). WMO State of the Global Climate 2025. WMO - 1360, World Meteorological Organization, https://reliefweb.int/report/world/state-global-climate-2025
5. Centre for Research on the Epidemiology of Disasters (CRED) & National Disaster Reduction Center of China (NDRCC). (2024). Global Natural Disaster Assessment Report. EM-DAT Insights. - https://www.preventionweb.net/media/84163/download?startDownload=20260707
6. Sun, Y., Zhang, L., Wang, J. (2024). Frameworks and challenges in digital twin-driven disaster risk management: A systematic review. International Journal of Disaster Risk Reduction. - https://www.sciencedirect.com/science/article/abs/pii/S2212420924000748?via%3Dihub
7. Rezaei, M., Terzi, S., Amaratunga, D. (2025). Intelligence-driven disaster management for smart urban resilience. Sustainable Cities and Society. https://doi.org/10.1016/j.scs.2024.105820
8. Semeraro, C., Lezoche, M., Panetto, H., Dassisti, M. (2021). Digital twin paradigm: A systematic literature review. Computers in Industry, 130, https://www.sciencedirect.com/science/article/abs/pii/S0166361521000762?via%3Dihub
9. Kritzinger, W., Karner, M., Traar, G., Henjes, J., Sihn, W. (2018). Digital Twin in manufacturing: A categorical literature review and classification. https://www.sciencedirect.com/science/article/pii/S2405896318316021?via%3Dihub
10. Liu, M., Fang, S., Dong, H., Xu, C. (2021). Review of digital twin about concepts, technologies, and industrial applications. Journal of Manufacturing Systems. https://www.sciencedirect.com/science/article/abs/pii/S0278612520301072?via%3Dihub
11. Rasheed, A., San, O., Kvamsdal, T. (2020). Digital Twin: Values, Challenges and Enablers from a Modeling Perspective. https://ieeexplore.ieee.org/abstract/document/8972429
12. Ford, D. N., Wolf, C. M. (2020). Smart Cities with Digital Twins: Looking Beyond Technology. https://www.researchgate.net/publication/342603052_Smart_Cities_with_Digital_Twin_Systems_for_Disaster_Management
13. Schrotter, G., Hürzeler, C. (2020). The Digital Twin of the City of Zurich as a Tool for Sustainable Urban Development. https://www.researchgate.net/publication/339039678_The_Digital_Twin_of_the_City_of_Zurich_for_Urban_Planning
14. GATE Institute (The Big Data for Smart Society Institute), Sofia, https://www.gate-ai.eu

Specific Features of Enterprise Liquidity Analysis

Bojidar Nachev

University of National and World Economy, Sofia, Bulgaria
bojiddar@unwe.bg

Abstract. Maintaining liquidity and solvency is one of the most important issues in the course of business of the enterprises. Enterprise liquidity represents its ability, at any given moment and without delay, to settle its obligations to its suppliers, creditors, employees and the State Budget with respect to taxes and social security contributions due on employees' remuneration. In the context of market economy conditions, enterprises organize their activities in accordance with the objectives of achieving maximum profits with minimum investment of resources, increasing profitability and maintaining adequate liquidity. By generating higher profitability, enterprises are able to expand the scope of their operations while maintaining or only minimally increasing their level of expenses. Only in this way, can they remain competitive on the market, increase their profits, and meet their current obligations to creditors through available cash resources. Resolving the "liquidity-profitability" dilemma links enterprises business to the development of liquidity risk. High liquidity results in reduced enterprise profitability, whereas low liquidity leads to indebtedness. The lack of liquidity necessitates the use of borrowings with uncertain interest rates, which in its turn increases liquidity risk and financial expenses of enterprises while reducing their profitability. An enterprise can be both liquid and profitable only if it makes effective investments in financial assets and instruments that increase the enterprise's rate of profit. The object of this paper is the specific features of enterprise liquidity analysis, while the subject is the opportunities for enterprises to manage liquidity risk.

Key words: Liquidity, liquidity risk, rate of return, solvency, enterprises.

1. Characteristics of enterprise liquidity

Liquidity is the ability of the enterprise to convert its non-current (fixed and long-term) tangible, intangible and financial assets, as well as its current (short-term and short) tangible and financial assets, into available cash with minimal losses. Enterprise liquidity represents its ability, at any given moment and without delay, to settle its obligations to its suppliers, creditors, employees and the State Budget with respect to taxes and social security contributions due on employees' remuneration. Furthermore, "...balance sheet liquidity should be understood as the degree to which the enterprise's current liabilities (payables) are covered by its owned and controlled current assets (short-term), whereby the period required to transform these assets into cash should correspond to the period for settling current liabilities." [1].

In the context of market economy conditions, enterprises organize their activities in accordance with the objectives of achieving maximum profits with minimum investment of resources, increasing profitability and maintaining adequate liquidity. By generating higher profitability, enterprises are able to expand the scope of their operations while maintaining or only minimally increasing their level of expenses. Only in this way, can they remain competitive on the market, increase their profits, and meet their current obligations to creditors through available cash resources.

Enterprise liquidity is characterized by the following major features:

A) It is the ability of enterprises to settle their obligations on time. To achieve this, enterprises must maintain sufficient cash to meet their current obligations to employees, suppliers, credit institutions, the state, social security institutions, and other creditors in timely manner;

B) It is the ability to fulfill the commitments undertaken by enterprises in timely manner, at reasonable cost, and with minimum liquidity risk;

C) It is the ability to settle current obligations in the event of large number of claims being simultaneously submitted by creditors and under conditions of high probability of liquidity risk. In such situation, enterprises should maintain liquid assets or be able to obtain liquidity quickly by borrowing external funds;

D) It is the ability to settle obligations not only to current but also to future counterparties and creditors.

In order to maintain their liquidity, enterprises should possess available cash, highly liquid short-term financial assets and financial instruments that can be readily traded on the stock market, or, when liquidity is needed, be able to obtain loans promptly from financial and credit institutions. "A prerequisite for maintaining liquidity is the synchronization by ... (enterprises – author's note) of incoming cash flows from all possible sources with outgoing cash flows arising from the settlement of any liabilities of ... (enterprises – author's note) to all their counterparties." [3].

Information for the purposes of enterprise liquidity analysis can be obtained from the Statement of Cash Flows and the Statement of Financial Position, respectively the Balance Sheet, prepared in accordance with the accounting framework adopted by the enterprises – the International Accounting Standards/ International Financial Reporting Standards (IAS/IFRS), respectively the National Accounting Standards (NAS).

2. Liquidity risk

Liquidity risk constitutes a serious threat to enterprises. It causes temporary or permanent loss of liquidity, as a result of which enterprises are unable to meet their obligations to employees, suppliers, credit institutions, the state, social security institutions and other creditors. This risk is closely related to interest rate risk, which results in future changes in interest rates that reduce enterprises' profits and lead to decline in interest income due to occurred market fluctuations in the interest rates of debt securities. Interest rate risk would reduce enterprise liquidity. The change in interest income of enterprises is determined by the structure of their investment and operating portfolios of financial assets and financial instruments, as well as by the direction of change in interest rate levels. These two risks are closely interconnected. They exert significant influence on both the liquidity and the profitability of enterprises. Interest rate is affected by factors that determine the condition of the money market, such as: - the average rate of profit;

- the supply and demand of financial assets and financial instruments on regulated stock markets;

- the government economic policy, etc.

Enterprises should maintain adequate level of liquidity that does not impede the enhancement of their profitability. If enterprises maintain excessively high liquidity, this would be at the expense of lower profitability. If they maintain excessively low liquidity, they would create the need to secure liquid funds in the form of borrowings and to operate in uncertain interest rate environment, thereby increasing enterprises' interest expenses and reducing their profitability. Therefore, enterprises should avoid low level of liquidity in their operations in order to prevent liquidity crises, insolvency and over-indebtedness.

3. Liquidity crisis in enterprise operations

A liquidity crisis in the operations of enterprises may arise for the following reasons:

- inability to settle current liabilities due to lack of available cash and cash equivalents;
- deteriorated financial performance resulting from incurred losses;
- failure to comply with agreed deadlines and significant delays in settling current liabilities;
- occurrence of unexpected payment obligations as a result of large number of claims submitted by creditors simultaneously;
- occurrence of global crisis (political, economic, health-related, such as the COVID-19 pandemic, etc.);
- occurrence of global energy crisis resulting from the wars involving Russia and Ukraine, USA and Iran, and Israel, Palestine and Lebanon, which have led to significant increase in fuel prices. Rising fuel prices cause increases in the prices of all raw materials, supplies, goods and services, leading enterprises to experience liquidity crises and contributing to rising inflation on global scale;
- increases in interest rates on loans obtained from financial and credit institutions as a consequence of rising inflation, etc.

4. Undertaking measures to address a liquidity crisis

The possible measures that enterprises should undertake to address a liquidity crisis include the following:

- renegotiating the deadlines for settling current liabilities with creditors;
- restructuring liabilities into long-term debt;
- converting debt owed to creditors into fixed capital;
- obtaining short-term borrowings from financial and credit institutions;
- selling owned financial assets and financial instruments on regulated stock markets;
- securing additional equity reserves;
- improving the collectability of receivables from customers through receivables insurance, sending notarized payment notices, freezing their bank accounts by court orders, selling uncollectible or difficult-to-collect receivables from customers through assignment to debt collection agencies;
- limiting the acquisition of fixed tangible and intangible assets;
- possibly quickly disposing of owned real estate, land and plots of land;
- increasing the sale prices of finished goods, merchandise and services offered for sale;
- increasing the fixed capital through a second issue of treasury shares;
- issuing corporate bonds
- obtaining additional financing of operations from a parent company or from a subsidiary, associate or joint venture;
- securing additional financial support by obtaining cash from the state in the form of subsidies or government grants and financial assistance;
- negotiating financial support agreements with other business counterparties;
- effectively selling other highly liquid assets – raw materials, supplies, merchandise, finished goods;
- introducing temporary prohibition on the distribution and payment of dividends and interests;
- reorganizing administrative and management expenses in order to limit non-essential expenditures;

- reorganizing the overall activities of the enterprise in the event of occurrence of liquidity crisis by increasing the "production volume, production capacity, capacity utilization by years, labor productivity:

✓ resource provision (key resources, markets, suppliers, prices..., etc.);

✓ sales and market risk management – market demand for the respective products, competitors, market shares, forecasts of changes in market demand, projected sales by years, target market share, market prices and forecasts of changes in market prices, marketing strategy, distributors and customers, etc." [4].

5. Liquidity indicators

Liquidity indicators are quantitative characteristics and that measure the enterprise's potential to meet its current liabilities using its available current assets. Liquidity ratios equal to or greater than 1 (one) indicate that the enterprise is able to meet its current liabilities with its available current assets.

"For the analysis and assessment of enterprise's liquidity, both absolute and relative indicators may be used, the values of which are determined on the basis of data from the balance sheet. In addition to the reporting date, for the purposes of liquidity analysis and management, average liquidity over the period under review may also be calculated by using data from the quarterly and monthly balance sheets, the trial balance, or the accounting records for the individual categories of short-term assets and short-term liabilities of the enterprise." [2].

A) Total liquidity ratio – it represents the ratio of the sum of current (short-term) assets to the sum of short-term liabilities of the enterprise. This ratio indicates that the enterprise fully covers its short-term liabilities with its available current (short-term) assets;

B) Quick liquidity ratio – it represents the ratio of the sum of short-term receivables (excluding difficult-to-collect and uncollectible receivables), short-term financial assets and cash to the short-term liabilities of the enterprise;

C) Immediate liquidity ratio – it represents the ratio of the sum of short-term financial assets and cash to the short-term liabilities of the enterprise;

D) Absolute liquidity ratio – it represents the ratio of available cash to the current liabilities of the enterprise. This indicator provides information on the most liquid assets and their amount that are immediately available to the enterprises and through which they meet their liabilities to creditors. If the value of this ratio is below 1 (one), this indicates that the enterprise has low liquidity and is unable to meet and satisfy the due claims of its creditors.

Where an enterprise does not meet the requirements to liquidity or has entered into prolonged liquidity crisis, it may become insolvent, over-indebted, insolvency proceedings may be initiated against it, and ultimately it may be declared bankrupt and liquidated.

6. Conclusion

Resolving the "liquidity-profitability" dilemma links enterprises business to the development of liquidity risk. High liquidity results in reduced enterprise profitability, whereas low liquidity leads to indebtedness. The lack of liquidity necessitates the use of borrowings with uncertain interest rates, which in its turn increases liquidity risk and financial expenses of enterprises while reducing their profitability. An enterprise can be both liquid and profitable only if it makes effective investments in financial assets and instruments that increase the enterprise's rate of profit. In this regard, D. Feschiyan and R. Andasarova state that "With the adoption of IFRS 9 Financial Instruments ... a new approach to the classification of financial assets was introduced, placing strong emphasis on the contemporary expected credit loss

recognition model. The newly introduced accounting framework for the recognition of credit risk impairment is a key instrument that directly affects the levels of regulatory capital ..." [5].

Enterprises should maintain an optimal level of liquidity that is adequate to the amount of their liabilities. In this way, liquidity and profitability come into direct conflict, as they are inversely related. Enterprise profitability, i.e. their rate of return, is often difficult to reconcile with the required level of liquidity. Enterprises' operations can be both liquid and profitable only if they efficiently invest in financial assets and financial instruments. Only in this case can enterprises increase their rate of profit without impairing their liquidity.

References

1. Ivanova, R., (2015) Aspects of Liquidity Analysis of Industrial Enterprises in Bulgaria, Report Collection of the National Scientific Conference on "Perspectives for and Challenges to Contemporary Management", UNWE, S., 2015, p. 169;
2. Ivanova, R., (2020), Analysis of Enterprise's Asset Profitability (a case study of industrial enterprises), Publishing Complex UNWE, S, 2020, p. 132, ISBN 978-619-232-325-7.
3. Natchkova, M., (2020), Relevant Issues of the Enterprise Liquidity Analysis, Report Collection of the 18th International Scientific Conference on "Management and Engineering'20", Technical University – Sofia, 2020, Sozopol, Republic of Bulgaria, p. 590;
4. Todorov, L., (2018), Investment Analysis and Management in Non-Financial Enterprises, Yearbook of the Institute of Certified Public Accountants (ICPA), 2018, p. 6], ISSN 1313-2148;
5. Feschiyan, D., Andasarova, R., (2020), Quality of Assets and Regulatory Capital – the New Challenge faced by the Bank Management, Collection of Reports from the Anniversary International Scientific Conference on the occasion of the 100th Anniversary of the Accounting and Analysis Department: "Contributions of Accounting to the Economic Science", Publishing House UNWE, S., 2020, p. 69.

Challenges to the Natural Disaster Protection System in the Republic of Bulgaria

Stefan Stefanov, Ekaterina Bogomilova

University of National and World Economy, Sofia, Bulgaria
stefan.stefanov@unwe.bg, ebogomilova@unwe.bg

Abstract. Natural disasters pose a significant risk to the life and health of the population, infrastructure, environment and economic development. The Republic of Bulgaria is exposed to various natural hazards, including earthquakes, floods, forest fires, landslides, droughts and extreme weather events, which necessitates the constant development and adaptation of the population protection system. The purpose of the study is to analyze the natural disaster protection system in the Republic of Bulgaria and to identify the main challenges to its effective functioning. The study is based on a documentary and comparative-analytical review of regulatory acts, national strategic documents and official documents of DG Fire Safety and Civil Protection, the European Commission and the United Nations Office for Disaster Risk Reduction (UNDRR) .The analysis highlights the importance of inter-institutional coordination, resource and technical provision, population preparation, early warning and the resilience of critical infrastructure. On this basis, measures and recommendations have been formulated, aimed at developing prevention, information exchange, technological capabilities and response readiness. The main conclusion is that increasing efficiency requires an integrated and adaptive management model that links risk assessment with planning, preparation, response and recovery.

Keywords: Natural disasters, Population protection, Risk management, Resilience, Prevention, early warning

1. Introduction

Natural disasters are among the major threats to the life and health of the population, the sustainable functioning of society and the economy. Despite the development of science and technology, earthquakes, floods, forest fires, landslides and extreme weather events can cause human and material losses, disruptions in critical infrastructure and long-term consequences for the affected territories (UNDRR, 2015; UNDRR, 2017b).

The need for effective protection of the population is becoming increasingly important in the context of a changing natural environment. International and European frameworks emphasize the assessment of existing and emerging risks, prevention, preparedness and resilience as interrelated strands of disaster management policy (UNDRR, 2015; European Commission, 2023).

Effective protection of the population is not limited to actions immediately before, during and after the disaster. It includes risk identification and assessment, prevention, early warning, preparation of institutions and the population, planning, resource provision, coordination and recovery (UNDRR, 2015; Council of Ministers of the Republic of Bulgaria, 2018). The complex nature of the activities requires periodic assessment of existing mechanisms and their timely improvement.

A significant challenge is the effective interaction between state and local authorities, civil protection services, health and emergency structures, voluntary groups, organizations and citizens. In the event of a large-scale natural disaster, timely exchange of information, proper allocation of responsibilities and coordination of actions are key to limiting the negative consequences (UNDRR, 2017a).

In this regard, modern technologies for monitoring, forecasting and early warning, as well as public awareness and preparedness, are of particular importance. The European disaster resilience goals identify risk assessment, public preparedness, early warning, response and the resilience of the civil protection system as mutually reinforcing areas (European Commission, 2023).

The purpose of the study is to analyze the existing natural disaster protection system in the Republic of Bulgaria, to identify the main challenges to its effective functioning and to formulate measures and recommendations to increase its effectiveness and sustainability.

The research thesis is that the effectiveness of the natural disaster protection system in the Republic of Bulgaria is not determined solely by the existence of a regulatory and institutional framework, but to a significant extent depends on the practical interaction between risk assessment, inter-institutional coordination, resource and technical provision, early warning, population preparation and the resilience of critical infrastructure. In this regard, increasing the effectiveness of the system requires interlinking prevention, preparation, response and recovery in disaster risk management.

To achieve the set goal, the following main tasks are set:

1. To systematize the main natural hazards and the associated risks for Bulgaria;
2. To analyze the regulatory and institutional framework for protection against natural disasters;
3. To identify the main challenges to the effective functioning of the system;
4. To assess the importance of prevention, early warning, resource provision and population preparation;
5. To analyze the role of inter-institutional coordination and modern technological solutions;
6. To formulate measures and recommendations to increase the efficiency and sustainability of the system.

2. Research methodology

The study is analytical in nature and uses documentary analysis, a comparative-analytical approach and a synthesis of normative, strategic and institutional sources. The Disaster Protection Act (DPA), national strategic and program documents, official materials of the National Disaster Risk Reduction and Management Committee, the Sendai Framework and the European Disaster Resilience Goals were analyzed. The selection of sources is limited to official normative, institutional and international documents directly related to the organization, prevention, preparation and management of natural disaster risk in the Republic of Bulgaria (UNDRR, 2015; Disaster Protection Act, 2006)-(UNDRR, 2025).

The documents were analyzed according to main thematic areas arising from the purpose and objectives of the study : normative and institutional organization of disaster protection, risk assessment and prevention, inter-institutional coordination, resource and technical provision, early warning, population preparation and resilience of critical infrastructure.

The comparative-analytical approach is used to compare the main principles and priorities in the national regulatory and strategic documents with the main guidelines for risk management and reduction set out in international and European documents. Based on the

analysis, the main challenges facing the functioning of the systemare identified and measures and recommendations are proposed to increase its effectiveness.

The limitation of the study is that it does not include a primary empirical study or quantitative measurement of the effectiveness of individual structures. Therefore, the conclusions and recommendations are analytical and application-oriented in nature and are based on the content of the considered normative, strategic and institutional sources.

3. Theoretical and normative foundations of population protection in natural disasters

Natural disasters can be considered as the result of the manifestation of hazardous natural processes and their interaction with exposed and vulnerable people, assets and systems. In international terminology, a "hazard" is a process or phenomenon that can cause loss of life, injury to health, material damage, socio-economic disruption or environmental destruction (UNDRR, 2017b).

Natural hazards vary in origin, mechanism, territorial scope, duration and impact. The international classification includes geological, hydrological, meteorological, environmental, biological and other types of hazards. Earthquakes and landslides are associated with geological processes, while floods, storms, droughts and temperature increases are associated with hydrological and meteorological processes (UNDRR, 2017b).

In Bulgarian conditions, among the natural hazards that can lead to disaster situations, earthquakes, floods, forest fires, strong winds and storms, snowfalls, icing and prolonged droughts stand out (GDPBZN, b.d.-4).

The modern approach to risk management distinguishes between hazard and disaster. The occurrence of a hazard does not automatically lead to a disaster; the consequences depend on exposure, vulnerability and available coping capacity. The Sendai Framework therefore places emphasis on understanding and reducing risk in advance, rather than solely on responding after an event has occurred (UNDRR, 2015; UNDRR, 2017b).

Major natural hazards can cause disruptions to the normal functioning of society, damage to buildings and infrastructure, disruption of transport and communication links, economic losses and human casualties. The severity of the consequences is a function not only of the characteristics of the hazard phenomenon, but also of the exposure and vulnerability of the affected systems. (UNDRR, 2015; UNDRR, 2017b)

In the Republic of Bulgaria, protection against natural disasters is mainly regulated by the Disaster Protection Act (DPA). It regulates public relations related to disaster protection, as well as the powers and interaction of competent authorities, legal entities and citizens (Disaster Protection Act, 2006).

The understanding that population protection should be a continuous process involving prevention, preparation, response and recovery is essential for the development of the system. This moves from a predominantly reactive model to one that prioritizes early identification of hazards, risk assessment, building response capabilities and increasing societal resilience.

This approach is also reflected in the National Strategy for Disaster Risk Reduction 2018-2030, adopted by Decision No. 505 of the Council of Ministers of 19 July 2018. The Strategy sets out the long-term framework for reducing existing risks, preventing new risks, increasing preparedness and response capabilities and improving recovery (Council of Ministers of the Republic of Bulgaria, 2018).

At the international level, the Sendai Framework sets out four priorities: understanding risk and resilience, strengthening risk management, investing in risk reduction and resilience,

and enhancing preparedness for effective response and recovery under the principle of "Build Back Better." (UNDRR, 2015).

Therefore, a natural disaster should be viewed not only as the result of the occurrence of a particular natural phenomenon, but as a complex process in which hazard, exposure, vulnerability and response capacity interact.

It is this understanding that is of particular importance for the present study, as it allows the challenges facing the population protection system to be considered not only from the perspective of disaster response, but also in terms of prevention, preparation and population resilience.

4. Regulatory and institutional framework for the protection of the population in natural disasters

The effective functioning of the system for protecting the population from natural disasters requires the presence of a clear regulatory and institutional framework that defines the powers, responsibilities and interaction between individual bodies and structures. In Bulgaria, this framework is built on laws and regulations, strategic and planning documents, as well as institutional mechanisms for prevention, preparation, response and recovery from disasters.

The main legal act in this area is the Disaster Protection Act (DPA). It regulates public relations related to the prevention and mitigation of the consequences of disasters, the organization and coordination of the actions of the competent authorities, as well as the participation of legal entities and individuals in disaster protection. The Act sets the legal basis for the functioning of the system and defines the main areas of activity related to prevention, preparation, response and recovery (Disaster Protection Act, 2006).

A significant place in the regulatory framework is occupied by the National Strategy for Disaster Risk Reduction 2018-20230. It is a long-term strategic document that defines the priorities and guidelines for disaster risk reduction in the country. The main goal is to create conditions for reducing existing risks, preventing and eliminating new ones, increasing preparedness and response capabilities, and improving recovery after a natural disaster (Council of Ministers of the Republic of Bulgaria, 2018).

The strategic framework is complemented by the National Program for Disaster Risk Reduction 2021-2025 and the annual plans for its implementation, through which the strategic objectives are specified in activities and measures (Council of Ministers of the Republic of Bulgaria, 2024; Council of Ministers of the Republic of Bulgaria, 2020).

Disaster protection plans are also of particular importance, through which the actions of the competent authorities are organized in the event of specific hazards. Planning allows for the determination of responsibilities, the necessary forces and means, the mechanisms for interaction and the ways of informing and warning the population in advance. This reduces the dependence on spontaneous decisions in crisis situations and creates the opportunity for a faster and more coordinated response.

The regulatory framework is supplemented by by-laws regulating the individual protection mechanisms. Since 2024, the establishment, maintenance, development and use of BG-ALERT is regulated by Ordinance No. 8121z-413 of March 29, 2024. (Ministry of Interior, 2024) The development of early warning systems is an essential element of modern protection. BG-ALERT uses Cell Broadcast to distribute warnings to compatible mobile devices in a specific geographical area and complements other warning and notification mechanisms (Ministry of Interior, 2024).

In institutional terms, the Ministry of Interior (MIA), through the General Directorate of Fire Safety and Population Protection (GDPBZN), has a key role. Along with operational activities, the GDPBZN carries out preventive activities, including informing the public and supporting preparation and planning. The preventive activities of the GDPBZN include informing the public about dangers and risks, supporting training and practical training, as well as activities related to planning and risk reduction (GDPBZN, p.d.-3). This confirms that the functions of the specialized structures are not limited to immediate response.

Local authorities are also of great importance. Municipalities are a key element of the system, since it is at the local level that a significant part of the practical measures for the protection of the population are implemented. Local authorities participate in the planning and organization of evacuation, informing the population, providing resources and coordinating actions on the territory of the respective municipality. This is of particular importance, since the consequences of natural disasters manifest themselves specifically on a certain territory and require knowledge of local risks and peculiarities.

Other institutions and organizations, including healthcare structures, local government bodies and voluntary groups, also participate in the system. The GDPBZN maintains a register of voluntary groups, and their development is also the subject of national, strategic and program documents (GDPBZN, b.d.-1; Council of Ministers of the Republic of Bulgaria, 2021, 2022). An important feature of the institutional framework is the need for coordination between different levels of government. Natural disasters do not respect administrative borders and in larger-scale events, several municipalities and districts can be affected simultaneously. This requires effective exchange of information, a clear distribution of responsibilities and the ability to promptly transfer and concentrate resources where they are most needed.

Another essential element is the control over compliance with regulatory requirements. The PBZN authorities carry out inspections of certain sites and plans and coordinate the development plans with regard to preventive measures and the protection of disaster-prone territories (GDPBZN, b.d.-2).

From the analysis of the regulatory and institutional framework, it can be concluded that Bulgaria has a multi-level risk management system, including legislative, strategic, planning and operational mechanisms, as well as tools for early warning, prevention, control and voluntary participation. However, the presence of a regulatory and institutional framework is not sufficient in itself to guarantee high efficiency. The real ability of the system depends on the degree to which regulatory requirements are translated into practical action, on the available resources, technical support, preparation of the human population, the quality of coordination and the ability to adapt to new and changing risks. It is the difference between a regulatory organization and practical efficiency that constitutes an important point for the subsequent analysis of the challenges and opportunities for its improvement.

5. Challenges facing the natural disaster protection system

The disaster protection system operates in a dynamic risk environment, which requires adaptation of prevention, preparedness, response and recovery mechanisms. Practical effectiveness depends on the ability to understand risk, coordination, resource provision and readiness for action - areas corresponding to the four priorities of the Sendai Framework (UNDRR, 2015).

One of the main challenges is the multifactorial nature of risk. It is shaped by the interaction between hazard, exposure, vulnerability and coping capacities, and therefore its management requires the collection, updating and analysis of information (UNDRR, 2015; UNDRR, 2017b).

This problem is of immediate importance for Bulgaria. The National Strategy for Disaster Risk Reduction 2018-2030 places emphasis on reducing existing risks, preventing new ones, and increasing preparedness and response and recovery capabilities. (Council of Ministers of the Republic of Bulgaria, 2018)

A significant challenge is inter-institutional coordination. According to UNDRR, risk management encompasses the system of institutions, mechanisms, policy and legal frameworks and other organizational means to guide, coordinate and oversee risk reduction activities (UNDRR, 2017a).

In the event of a large-scale natural disaster, central and local authorities, fire and civil protection structures, health and rescue teams and voluntary groups may be involved simultaneously. Timely exchange of information and clear allocation of functions are a prerequisite for the effective use of available capabilities (UNDRR, 2015; UNDRR, 2017a).

The practical importance of coordination and advance preparedness is highlighted by the floods caused by intense rainfall in October 2025, which affected various regions of the country, including the districts of Burgas, Montana and Pernik. The complicated situation necessitated the declaration of a state of emergency in individual municipalities, the activation of the operational headquarters of the Ministry of Interior and the activation of the flood action plan. This example shows the need for prevention, risk assessment and inter-institutional coordination to be linked to the readiness for a timely response to rapidly developing natural disasters (Ministry of Interior, 2025b).

Another important aspect is the resource and technical provision of the disaster protection system. The ability to respond effectively is directly related to the availability and readiness of the necessary forces, technical means and organizational capacity. Modern risk management, however, implies that resources should be directed not only to immediate response, but also to preventive measures. The Sendai Framework defines investment in risk reduction and resilience as one of the four main priorities, linking such investments to increasing the economic, social, health and physical resilience of populations, communities and their assets (UNDRR, 2017a).

In Bulgaria, the need to maintain and develop capacities is enshrined in the National Strategy for Disaster Risk Reduction 2018-2030, the National Program for Disaster Risk Reduction 2021-2025 and the annual plans for its implementation. (Council of Ministers of the Republic of Bulgaria, 2018; Council of Ministers of the Republic of Bulgaria, 2024; Council of Ministers of the Republic of Bulgaria, 2020) This shows the need for resource provision to be considered as part of a systematic process of planning and capacity building, and not solely as a response to a disaster that has already occurred.

The preparation of the human factor is also of particular importance. The effective use of regulatory mechanisms, plans, technical means depends on the knowledge and abilities of the people who implement them. In the event of a natural disaster, decisions are often made under conditions of limited time, a high degree of uncertainty and the need for interaction between different structures. Therefore, preparation, training and practical playing of different scenarios are an essential element of building response capabilities. The Sendai Framework emphasizes the readiness for an effective response and the availability of necessary capabilities at different levels of government (UNDRR, 2015).

Along with the professional training of competent structures, an important challenge is the preparedness of the population for action in the event of disasters. Citizens should not be considered only as objects of protection, but also as participants in the risk reduction process. Understanding the existing hazards and how to behave in them is directly related to the first priority of the Sendai Framework - understanding disaster risk (UNDRR, 2015). Raising

awareness and building a culture of prevention can support the correct behavior when receiving a warning, the need for evacuation or the occurrence of an immediate danger.

The main challenges include translating risk assessment into practical preventive actions. The availability of information on hazards and vulnerable areas has real value only when this information is used in planning and decision-making. Understanding risk should serve as a basis for assessment, prevention, mitigation, preparation and response (UNDRR, 2015). This implies that the results of the risk assessment should be linked to the planning of protection, development and maintenance of infrastructure, the determination of the necessary resources and the preparation of competent structures.

The analysis shows that the challenges facing the disaster protection system are complex and highly interconnected. Insufficient risk information can lead to misprioritization, resource constraints, reduced operational capabilities, and insufficient coordination to hinder the effective use of available resources. This is why a modern approach requires integrated management, in which risk awareness, institutional coordination, investments in resilience and preparedness are seen as complementary components in a single system (UNDRR, 2015; UNDRR, 2017a).

The identified challenges are examined in their interrelationships within the study. The normative, institutional and practical analysis create a basis for determining priority measures aimed at increasing the resilience and effectiveness of the natural disaster protection system.

6. Measures and recommendations for increasing the effectiveness of the natural disaster protection system in the Republic of Bulgaria

The analysis of the regulatory and institutional framework, the main challenges and practical phenomena of natural disasters in the Republic of Bulgaria show that increasing the effectiveness of the protection system requires a complex of interrelated measures. They should be aimed not only at increasing the response capabilities to an already occurring disaster, but also at improving prevention, risk assessment, early warning, population preparation and recovery. Such an approach corresponds to both the National Strategy for Disaster Risk Reduction 2018-2030 and the priorities of the Sendai Framework for Disaster Risk Reduction 2015-2030 (Council of Ministers of the Republic of Bulgaria, 2018; UNDRR, 2015).

First, there is a need to improve the processes of risk identification, assessment and monitoring. Effective prevention requires that decisions are based on up-to-date information on hazards, the vulnerability of populations and infrastructure, and the possible consequences under different scenarios. This implies periodic updating of risk assessments, maps of threatened areas and protection plans, as well as the wider use of geographic information systems, remote sensing and other digital tools. The Sendai Framework sets the understanding of risk as the first priority and emphasizes that information on hazards, exposure, vulnerability and capacity should serve as the basis for prevention, preparation and preparedness (UNDRR, 2015).

For Bulgaria, improving the link between risk assessment and practical planning is of particular importance. Information about risk areas should be used more actively in spatial planning, the construction and maintenance of infrastructure, the determination of necessary protective facilities and the planning of response resources. In this way, prevention can be aimed at limiting vulnerability even when a crisis situation arises, and not only at reducing the consequences after it has occurred.

The second main direction is the enhancement of inter-institutional coordination and information exchange. Natural disasters usually require interaction between central and local government bodies, the National Disaster Management Authority, health structures, voluntary groups and other organizations. The effectiveness of their interaction depends on the presence of pre-defined mechanisms for information exchange, a clear distribution of functions and a unified assessment of the operational situation. Strengthening risk management and coordination between actors is also identified as an independent priority in the Sendai Framework (UNDRR, 2015).

In this regard, it is appropriate to develop an integrated information environment for disaster risk management, allowing the exchange of up-to-date data at national, regional and municipal levels. Such a system could bring together information on events that have occurred, available forces and resources, risky infrastructure sites, evacuation routes and other data necessary for operational management. This would reduce the time for receiving and processing information and would support decision-making in the event of a crisis situation.

The third direction is the improvement of the resource and technical provision of the specialized structures. The effectiveness of the response depends to a significant extent on the availability of sufficient and technically sound means, allowing work in various types of natural disasters. The official strategic framework of the GDPBZN also includes a Program for the modernization and technical provision of the territorial units for fire safety and population protection, which indicates the need to maintain and renew the material and technical capacity (GDPBZN, b.d.-5).

When determining resource priorities, it is necessary to take into account not only the current state of technology, but also the change in risk areas. Large-scale fires, floods and extreme weather events may require the simultaneous use of a significant number of teams and specialized assets in different parts of the country. This requires periodic assessment of regional capabilities and the development of scenarios for the timely redirection and concentration of necessary resources.

The need for sufficient resource and operational capacity was also evident during the forest fires in the area of the village of Ilindentsi, Strumian Municipality in the summer of 2025. The prolonged actions to control it required the participation of the employees of the State Fire Service for Forest Fire Prevention and Control, forest officers and volunteers, as well as the use of air assets. The case showed the importance of the possibility of timely concentration of additional forces and equipment and for coordinating actions in large-scale operations (Ministry of Internal Affairs, 2025a).

The fourth direction is the improvement of early warning and public notification systems. The introduction of BG-ALERT represents a significant step in the development of national capabilities for disseminating warnings. However, the effectiveness of the system depends not only on the technical ability to reach citizens, but also on the timeliness, clarity and specificity of the messages, as well as on the knowledge of the population of the necessary actions after receiving them. The regulatory basis and conditions for the use of BG-ALERT are determined by Regulation No. 8121z-413 of March 29, 2024 (Ministry of Interior, 2024).

The European Disaster Resilience Goals also identify improving early warning as one of the five main areas, along with risk assessment, population preparation, response and resilience of the civil protection system. (European Commission, 2023) Therefore, the development of BG-ALERT should be accompanied by periodic testing, analysis of the effectiveness of the warnings sent and information campaigns on their importance and the correct behavior in different types of hazards.

The fifth area is raising the population and building a culture of prevention. Even a well-functioning institutional system can be hampered if the population does not know the

risks and does not know how to react to warnings, evacuations and imminent danger. The European Commission sets increasing the risk awareness and preparedness of the population as a stand-alone objective for resilience to natural disasters (European Commission, 2023).

A practical measure in this area would be to conduct regular information and training campaigns tailored to the main risks for specific regions. In areas with an increased risk of floods, forest fires or earthquakes, the training of the population should include clear algorithms of behavior, evacuation, first aid and use of official information channels. Of particular importance is the inclusion of such training in the educational system and the conduct of practical classes and exercises.

The sixth direction is the active development and use of voluntary formations. The Strategy for the Development of Voluntary Formations 2021-2030 was adopted by Decision No. 740 of the Council of Ministers of October 28, 2021, the National Program for the period 2022-2026 - by Decision No. 758 of October 12, 2022. (Council of Ministers of the Republic of Bulgaria, 2021, 2022) GDPBZN maintains specialized information and programs for their training (GDPBZN, b.d.-1; Council of Ministers of the Republic of Bulgaria, 2021, 2022).

In this area, it is advisable to encourage the creation and maintenance of effective voluntary formations in municipalities with higher natural risk, as well as to ensure regular training, equipment and participation in joint exercises with professional structures. Voluntary formations should not replace professional rescue services, but should increase local capacity and initial response capability.

The seventh area is to increase the resilience of critical infrastructure and essential public services. Experience with floods and fires shows that a natural disaster can cause secondary consequences through disruption of electricity, water supply, transport links and communication. Therefore, the risk assessment should cover the interdependencies between infrastructure systems and identify the sites whose disruption would have the most serious consequences for the population.

For critical facilities, business continuity and recovery plans should be developed and periodically updated. Alternative communication channels, backup power supplies, temporary water supply options, and alternative transportation routes should be identified periodically. Investing in structural and non-structural resilience measures is among the core principles of the Sendai Framework (UNDRR, 2015).

It is also necessary to systematically learn from past disasters. After major crisis events, a standardized analysis should be carried out to assess the timeliness of warning, coordination, resources used, evacuation actions, communication with the population and recovery activities. The results of such analyses should lead to specific changes in plans, preparations and resource allocation.

Based on the analysis carried out, it can be summarized that increasing the effectiveness of the natural disaster protection system in Bulgaria requires increasing preventive activities in risk management. Important in this regard are timely identification and assessment of risk, development of local and national capacity, technical support, improvement of the information model and increasing the preparedness of the population.

The proposed measures should be considered as interrelated. The effectiveness of early warning systems is related to the willingness of the population to put them into practice in planning and preparation. The integration of these components creates an opportunity to increase the resilience of the system and limit the human, material, economic, and environmental consequences of future natural disasters.

The scientific and applied contribution of the study can be linked to the systematization of basic normative, institutional and organizational factors that are important for the effectiveness of protection against natural disasters in Bulgaria, as well as to the delineation of

specific directions for its improvement. The conclusions drawn can be used as a basis for subsequent research and the development of indicators for assessing the preparedness and resilience of the system.

7. Conclusion

Natural disasters represent a complex challenge to the security of the population, infrastructure, economy and environment. The analysis confirms the need for the disaster protection system in the Republic of Bulgaria to develop as a flexible and integrated system combining prevention, preparation, response and recovery. Bulgaria has a well-established regulatory and institutional framework for risk management and disaster protection, which creates the basis and organization of activities in this area.

The regulatory and institutional basis of protection against natural disasters in the Republic of Bulgaria is determined by the Law on Protection against Natural Disasters (ZDB) , the National Strategy for Disaster Risk Reduction 2018-2030 and the related planning and program documents. The practical implementation of preventive and operational activities is based on the interaction between the competent state and local authorities, with the General Directorate for Fire Safety and Population Protection (GDPBZN) , together with the other structures of the Unified Rescue System and the voluntary formations, playing a key role in this process.

The existence of a regulatory and organizational framework alone does not guarantee the effectiveness of disaster protection, which depends on a number of interrelated factors. The analysis highlights the need to improve inter-institutional coordination and information exchange, maintain and develop human and technical capacity, increase the effectiveness of early warning systems, improve population preparedness, and strengthen the resilience of critical infrastructure.

The analysis of the normative, strategic and institutional sources emphasizes the need for practical consideration of the identified challenges. Of particular importance in this regard are the timely mobilization and redirection of resources, coordination between institutions at different administrative levels, timely information of the population and maintenance of the resilience of critical infrastructure. These aspects should be reflected both in the planning and preparation processes and in the subsequent evaluation of the actions taken in the event of natural disasters.

On this basis, it can be concluded that the development of the disaster protection system should be aimed at strengthening prevention and increasing preparedness for action in the event of natural disasters. Of particular importance in this regard are the systematic identification and assessment of risk, the use of the results of these assessments in planning, the improvement of coordination between responsible structures, the development of technical capabilities and the improvement of early warning systems.

An important factor for the effectiveness of the disaster protection system is the readiness of human resources for action. Improving the professional skills of employees and volunteer formations should be combined with increasing awareness and readiness of the population for adequate behavior in case of natural disasters. This would support timely response and contribute to limiting the consequences in case of disaster situations.

The measures proposed in the study emphasize that increasing the effectiveness of the disaster protection system requires improving not only its individual components, but also the interaction between them. This implies the implementation of an integrated approach, uniting the development of the regulatory framework, institutional interaction, resource provision, technological capabilities, early warning systems, population preparation and the resilience of critical infrastructure.

The results of the analysis confirm the research thesis that the effectiveness of the disaster protection system is not determined solely by the presence of a regulatory and institutional framework, but depends on the practical interaction between its main components. The established interdependence between risk assessment, inter-institutional coordination, resource and technical provision, early warning, human factor preparation and the resilience of critical infrastructure emphasizes the need for their coordinated development. In this context, increasing the effectiveness of the system requires a consistent linking of prevention and risk assessment with planning, preparation, response and recovery.

In conclusion, it can be summarized that the sustainable development of the natural disaster protection system in Bulgaria is related to its ability to timely identify changing risks, to draw lessons from disasters that have occurred and to use the accumulated experience in planning and preparation. The consistent implementation of preventive and organizational measures, combined with improved coordination and increased response readiness, can contribute to reducing the vulnerability of society and limiting the consequences of future natural disasters.

Acknowledgments

This work has been carried out in the framework of the Project "Application of Artificial Intelligence for Sustainable Business Development in the Context of Climate Change", financially supported by the Science Fund of the University of National and World Economy, Bulgaria (Grant No. NID NI 35/2025/A).

References

1 European Commission. (2023). Commission recommendation of 8 February 2023 on Union disaster resilience goals (2023/C 56/01). Official Journal of the European Union, C 56, 1-11. https://eur-lex.europa.eu/legal-content/BG/ALL/?uri=CELEX:32023H0215(01)

2 United Nations Office for Disaster Risk Reduction (UNDRR). (2015). Sendai framework for disaster risk reduction 2015-2030. https://www.undrr.org/publication/sendai-framework-disaster-risk-reduction-2015-2030

3 United Nations Office for Disaster Risk Reduction (UNDRR). (2017a). Disaster risk governance. https://www.undrr.org/terminology/disaster-risk-governance

4 United Nations Office for Disaster Risk Reduction (UNDRR). (2017b). Hazard. https://www.undrr.org/terminology/hazard

5 United Nations Office for Disaster Risk Reduction (UNDRR). (2025). UNDRR strategic framework 2026–2030: Accelerated action for a safer, more resilient future. https://www.undrr.org/strategic-framework-2026-2030

6 General Directorate of Fire Safety and Population Protection (GDPBZN). (b.d.-1). Voluntary formations. https://www.mvr.bg/gdpbzn/дирекцията/дейности-на-гдпбзн/dobrovolni_form

7 General Directorate of Fire Safety and Population Protection (GDPBZN). (b.d.-2). Control over compliance with the Disaster Protection Act. https://www.mvr.bg/gdpbzn/дирекцията/дейности-на-гдпбзн/контрол-по-спазвань-на-закона-за-заштита-при-бедствия

8 General Directorate of Fire Safety and Population Protection (GDPBZN). (b.d.-3). Preventive activities. https://www.mvr.bg/gdpbzn/дирекцията/дейности-на-гдпбзн/pd

9 General Directorate of Fire Safety and Population Protection (GDPBZN). (b.d.-4). In other disaster situations. https://mvr.bg/gdpbzn/info-center/pravila-povedenie/pri-bedstvia

10 General Directorate of Fire Safety and Population Protection (GDPBZN). (b.d.-5). Program for modernization and technical provision of the territorial units for fire safety and population protection in the Republic of Bulgaria. https://mvr.bg/gdpbzn/дирекцията/нормативна-уредба/стратегически-документи

11 Disaster Protection Act. (2006). State Gazette, No. 102 of 19 December 2006 (with subsequent amendments and supplements). https://dv.parliament.bg/DVWeb/showMaterialDV.jsp?idMat=2971

12 Council of Ministers of the Republic of Bulgaria. (2018). National Strategy for Disaster Risk Reduction 2018–2030 (Decision No. 505 of 19 July 2018). https://mvr.bg/gdpbzn/дирекцията/нормативна-уредба/стратегически-документи
13 Council of Ministers of the Republic of Bulgaria. (2020). National Disaster Risk Reduction Program 2021–2025 (Decision No. 865 of 26 November 2020). https://mvr.bg/gdpbzn/дирекцията/нормативна-уредба/стратегически-документи
14 Council of Ministers of the Republic of Bulgaria. (2021). Strategy for the Development of Voluntary Formations for the Prevention or Management of Disasters, Fires and Other Emergency Situations in the Republic of Bulgaria for the Period 2021–2030 (Decision No. 740 of October 28, 2021). https://mvr.bg/gdpbzn/дирекцията/нормативна-уредба/стратегически-документи
15 Council of Ministers of the Republic of Bulgaria. (2022). National Program for the Development of Voluntary Formations for the Period 2022–2026 (Decision No. 758 of October 12, 2022). https://mvr.bg/gdpbzn/дирекцията/нормативна-уредба/стратегически-документи
16 Council of Ministers of the Republic of Bulgaria. (2024). Annual Plan for 2024 for the Implementation of the National Program for Disaster Risk Reduction 2021–2025 . https://mvr.bg/gdpbzn/дирекцията/нормативна-уредба/стратегически-документи
17 Ministry of Internal Affairs. (2024). Regulation No. 8121z-413 of March 29, 2024 on the procedure for the construction, maintenance, development and use of the BG-ALERT system for the dissemination of warning messages to the population. https://mvr.bg/gdpbzn/дирекцията/нормативна-уредба/наредби/naredba_bedstvia
18 Ministry of Internal Affairs. (2025a, August 4). Current information on the fire situation in the territory of Blagoevgrad district. https://mvr.bg/press/aktualna-informacija/aktualna-informacija/новини/83768
19 Ministry of Internal Affairs. (2025b, October 3). Intense rainfall and dynamic meteorological conditions affected the districts of Burgas, Montana and Pernik. https://mvr.bg/press/aktualna-informacija/aktualna-informacija/новини/86394

Some Aspects of the Connection between Artificial Intelligence, Sustainable Business and Climate Change

Milena Kunova, Katerina Doychinova

University of National and World Economy, Sofia, Bulgaria
milena.kunova@abv.bg, doichinova58@gmail.com

Abstract. This study examines the set of challenges, risks and opportunities related to digitalization and the implementation of artificial intelligence (AI) in the context of climate change and business sustainability. The focus is on the European Union, with the United States and China considered as reference examples of global trends, as they represent leading forces in the field of digital transformation of the economic business sector. In this way, the depth of regional analysis is combined with an international perspective. The report also covers the economic and educational aspects of digital transformation, analyses key scientific reports, presents examples of successful practices and critical moments related to cybersecurity and sustainability, and offers recommendations for integrating AI and digital technologies into the climate strategies of business and society. The study also cites regulatory examples. Steps for action are proposed, such as regulatory frameworks and balanced approaches. The need for sustainability, international cooperation and critical assessment of digitalization and AI is emphasized.

Keywords: Digitalization, risks, climate change, business sustainability, AI

1. Introduction

Digitalisation and artificial intelligence are becoming the leading drivers of transformation in the economy, education and sustainable development [4, 5]. The European Commission inspects the digitalisation as a key factor for increasing competitiveness and economic sustainability [6]. The Organisation for Economic Co-operation and Development (OECD) sees digitalisation as a driver of economic growth, but also as a source of new social inequalities. The United Nations Educational, Scientific and Cultural Organisation (UNESCO) has highlighted the role of digitalisation in education, especially during the pandemic [7, 8]. At the same time, the World Economic Forum has warned of growing digital dependencies and vulnerabilities [9, 10]. However, in recent years, AI has been playing an increasingly important role in addressing global challenges, including climate change, energy management and the green transformation of business. Digitalisation and AI offer new opportunities for business and education, but also pose new challenges, especially in the context of climate change. According to a 2025 World Economic Forum report [10], digital technologies have the potential to reduce global emissions by up to 15% by 2030.

Artificial Intelligence (AI) adds a new dimension, providing capabilities for forecasting climate events, optimising energy systems, and sustainable agriculture.

This study aims to analyse the effects and risks of digitalisation, including AI, with an emphasis on the need for sustainability and balanced development.

The study is based on a literature review, comparative analysis and case studies. The main sources are reports of international organisations (EU, OECD, UNESCO, World Bank, World Economic Forum), as well as scientific publications. The methodology includes:

Review of academic and institutional sources. Analysis of specific cases. Comparison of policies and approaches in the EU, the US and China. Identification of good practices and regulatory approaches in the field of AI and climate technologies [18].

2. Digital transformation of the economy and AI

In recent decades, digital technologies have transformed economic models. Fintech companies like Revolut and PayPal are introducing new business models based on mobile and cloud solutions. At the same time, AI is emerging as a critical technology for sustainable governance and ecological transition. In the European Union, over 60% of companies use cloud services. At the same time, the World Bank [9] emphasises that digital finance [12] also carries a risk of household indebtedness. As Dimitrov [2] notes, the knowledge economy in Bulgaria is also being formed under the influence of digitalization.

AI and Sustainable Business Management

- Predicting climate events – using machine learning and satellite imagery analysis (Google AI for Climate Change) [13, 14].
- Optimising energy grids – Tesla uses AI to manage batteries and predict demand [15].
- Agriculture – John Deere applies AI to analyse soils, implement precision irrigation, and reduce fertiliser use, making businesses more sustainable [16].

Energy costs and the carbon footprint of AI

The training of the AI models is creating high energy intensity. According to the journal Nature Communications, one large-scale AI model can generate carbon emissions comparable to tens of thousands of domestic flights.

The impending digitalization also poses huge risks for business. There are hundreds of examples of this, such as the global cyberattack of the WannaCry malware in 2017, which affected hundreds of thousands of computers and blocked hospitals and banks. In 2020, the attack on the American software development company SolarWinds compromised American government institutions. In 2021, another American company, Colonial Pipeline, fell victim to a malicious software attack that led to an interruption in fuel supplies in the United States. These examples are part of a multi-layered problem that shows how vulnerable economic systems are in conditions of digital dependence. This shows that business is at risk and urgent measures are needed to protect against such attacks. Figure 1 shows the level of digitalization by regions (EU, USA, China).

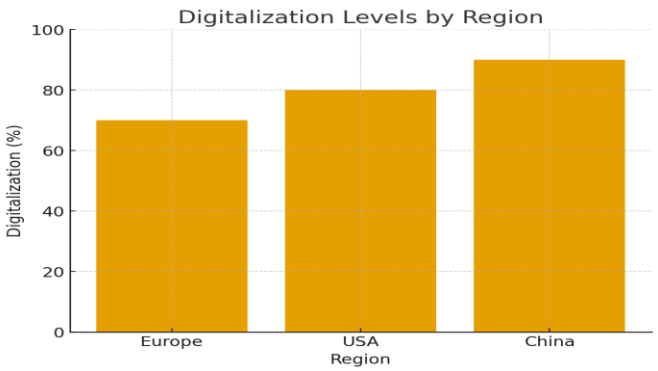


Fig. 1. Level of digitalization by region (EU, USA, China).

Digital Integration in Education

The COVID-19 pandemic has accelerated the digitalisation process in education. UNESCO reported that over 1.5 billion students worldwide have been forced to switch to online learning. Online platforms (Zoom, Microsoft Teams, and Moodle) have become the main means of communication and teaching. In Bulgaria, Shkolo.bg has become the central school management system. According to Georgieva [1], digital technologies are becoming an indispensable part of modern education.

However, risks include unequal access to the Internet, social isolation of students, screen fatigue, and reduced concentration. In 2020, UNICEF reports showed that children from low-income families were significantly disadvantaged due to a lack of devices. According to UNESCO [7], over 1.5 billion students worldwide were affected by school closures. The Organisation for Economic Cooperation and Development (OECD) emphasises that digital education provides flexibility, but at the same time deepens inequalities between students with different socio-economic statuses. In Bulgaria, the Ministry of Education and Science reported serious challenges in accessing devices and the Internet for about 10% of students. In the EU, special programs for digital inclusion were adopted, including financing laptops and tablets for socially vulnerable groups [17]. These consequences of the pandemic shook up business.

Critical Aspects and Vulnerabilities of Digitalization and AI

Digitalization and the role of AI bring many risks and limitations that can be summarised in the following categories:

- Dependence on infrastructure - lack of electricity or internet leads to paralysis of economic and educational processes.
- Cybersecurity - attacks on banks, companies and government institutions show the real threats.
- Social inequality - the digital divide deepens the differences between social groups and regions.
- Excessive automation - leads to the loss of critical thinking and traditional skills. In support of this thesis, Taiwan in 2015 adopted a law that limits the time that children can spend in front of electronic devices, providing for fines for parents who do not comply with the norms.

According to Sandia Zeverte-Rivza [11], the risks associated with cybersecurity and technology are most noticeable with digitalisation. Pavlov [3] focuses on cybersecurity issues and the need for national strategies.

Regulatory mechanisms and legal framework

There is a fragmentation of regulations and a lack of effective oversight. Despite the growing role of digitalisation and AI, the monitoring and regulation of security and ethical standards remain fragmented and often ineffective:

- At the international level, the UN, UNESCO and OECD formulate principles and recommendations, but these are not legally binding.
- In the European Union, the AI Act (2021–2024) and the Digital Services Act (2022) are the first attempts to create a legal framework, but their actual implementation is yet to come [17]. The European Union Agency for Cybersecurity (ENISA) [5] plays a coordinating role, but does not have sufficient sanctioning powers.
- In the US, there is no unified federal framework – the National Institute of Standards and Technology (NIST) is developing voluntary guidelines (e.g. AI Risk Management Framework, 2023), while individual states such as California are adopting their own

regulations for the protection of personal data (California Consumer Privacy Act, 2018).

- China has centralised laws on algorithms and online time for minors, but transparency and compliance with international standards remain a problem.
- Companies (Google, Microsoft, IBM, etc.) maintain internal codes of ethics and committees that are often formal and subordinate to business interests.
- Civil society and academic organisations play the role of a critical observer, but without real regulatory power.

These contrasts show the lack of global harmonisation and the difficulty of developing common international standards.

This fragmentation leads to chaos and a lack of effective oversight, which increases risks and creates gaps in global digital security and poses many risks for business and the economy as a whole. In the Bulgarian context, the problem is also discussed in the collection Digital Transformation and Education – Challenges and Opportunities (2020).

Examples of key regulatory documents include: Regulation (EU) 2022/2065 on Digital Services (Digital Services Act), which creates a framework for the liability of online platforms; and the proposal for a Regulation of the European Parliament and of the Council on Artificial Intelligence (AI Act) which defines risk categories for AI systems and provides mandatory requirements for high-risk applications. These documents are still being phased in, highlighting the dynamic and uncertain regulatory environment.

Table 1. Benefits and risks of digitalization and AI

Risk aspects	**Examples and consequences**
Infrastructure dependency	Without internet/power supply, basic economic and educational processes are blocked
Cybersecurity	Attacks: WannaCry 2017, SolarWinds 2020, Colonial Pipeline 2021, etc.
Social inequality	The digital divide between regions and social groups is deepening
Energy consumption	Model training requires resources, often from non-renewable sources

Outlined implementation actions

To reduce the risks of malicious digitalization and AI, the following steps are needed:

1. Development and implementation of regulatory frameworks – EU with the Digital Europe Programme, US through the National Institute of Standards and Technology (NIST) guidelines.
2. Balance between digital and backup systems – providing alternative solutions in case of network failure.
3. Investment in cybersecurity – establishing corporate ethics committees and protecting critical infrastructure.
4. Training and development of personnel: Preparing AI specialists for sustainable business.
5. Investment in sustainable AI technologies: Developing AI solutions that minimise energy consumption and carbon emissions.

6. International cooperation: Creating global standards and regulations for the ethical use of AI.
7. Integration of AI in climate strategies: Using AI to predict climate events and optimise resources for sustainable business.

According to the World Bank, digital transformation will only be successful if it is accompanied by social and regulatory safeguards.

3. Conclusion

Digitalisation is an inevitable process that fundamentally changes the economy, education and business. Despite the many advantages, it also creates serious risks – from cybersecurity to social inequalities. Historical examples show that excessive dependence on technology can lead to systemic and harmful failures.

It is necessary to build a balanced development model – one that combines innovation with sustainability and security. Regulations at the international and national levels, the active role of business and educational institutions, as well as the inclusion of civil society, are key to achieving sustainable digital transformation. The topic is dynamic and subject to development, and future research should focus on the creation of global standards and a balance between technology and the human factor. In addition to economic and technological aspects, it is extremely important to focus on human capital – skills, critical thinking and social responsibility. The lack of global standards and uniform regulations exacerbates the risks and requires international coordination. Future research should focus on the long-term social effects of digitalisation, including the consequences of mass online learning during COVID19 and their impact on equality, mental health and the quality of education. Only through a holistic and balanced approach can the digital transformation be sustainable and beneficial for society.

Artificial intelligence offers real solutions to address climate challenges and build sustainable businesses. At the same time, it creates risks – both environmental (due to high energy consumption), social and ethical. A balanced approach is needed: AI must be integrated responsibly, under strict regulatory control, for more sustainable business in the face of climate change. The approach to sustainability requires global coordination, educational transformation, a focus on human capital and an international ethical balance for business development.

Acknowledgments

This work has been carried out in the framework of the Project "Application of Artificial Intelligence for Sustainable Business Development in the Context of Climate Change", financially supported by the Science Fund of the University of National and World Economy, Bulgaria (Grant No. NID NI 35/2025/A).

References

1. Georgieva T. (2019). Information and Communication Technologies in Education. Sofia: St. Kliment Ohridski University.
2. Dimitrov M. (2020). Knowledge Economy and Digitalization. Sofia: UNWE
3. Pavlov, P. (2021). Cybersecurity and Risks in the Digital Economy. Economic Research, vol. 30, no. 4, pp. 55–70.
4. Collection of Papers (2020). Digital Transformation and Education - Challenges and Opportunities. Sofia: St. Kliment Ohridski University.
5. ENISA (2022). ENISA Threat Landscape 2022.
6. European Commission (2020). Shaping Europe's Digital Future.

7. UNESCO (2021). Education during COVID-19 and beyond.
8. UNICEF (2020). COVID-19: Are children able to continue learning during school closures?
9. World Bank (2022). Digital Development Report.
10. World Economic Forum (2020). The Global Risks Report 2020.
11. Zeverte-Rivza S. et al. (2024). Sustainability, vol. 16, no. 2, art. no. 524.
12. ArXiv (2022). The rise of digital finance: Financial inclusion or debt trap.
13. Nature Communications (2021). Energy Consumption of Machine Learning Models. Retrieved from https://www.nature.com/articles/s41598-021-86485-2
14. Google AI for Climate Change (2025). Retrieved from https://ai.google.com/research/
15. Tesla, Optimising Energy Consumption with AI (2025). Retrieved from https://www.tesla.com
16. John Deere Precision Agriculture (2025). Retrieved from https://www.deere.com
17. European Commission AI Act (2025). Retrieved from https://ec.europa.eu/digital-strategy/our-policies/artificial-intelligence_en
18. Climate Data Initiative, USA (2025). Retrieved from https://www.data.gov/climate/

Self-Service Systems in the Private and Public Sector: Adoption, Benefits, and Challenges

Petar Genov

University of National and World Economy, Sofia, Bulgaria
petar.genov@unwe.bg

Abstract. Self-service systems are now widely implemented globally. This technology provides a digitalized service that enables customers to resolve issues without staff intervention. This paper focuses on analyzing the service in both private and public sector as it focuses on adoption, benefits and challenges for both the consumer and the provider.

Keywords. Self-service systems, privacy, security, public sector, private sector.

1. Introduction

Self-service systems are now widely implemented globally. This technology provides a digitalized service that enables customers to resolve issues without staff intervention. Self-service technologies (SST) are utilized in retail, supermarkets, travel, hospitality, banking, and customer support. Studies suggest that "advancements in information technology have changed the way customers experience a service encounter and their relationship with service providers. Especially technology-based self-service channels have found their way into the 21st century service economy." [34] The expansive advancement of digital technologies has resulted in the widespread adoption of self-service systems in both private and public sectors. Technologies such as ATMs, e-government portals, self-service kiosks, and mobile applications have become integral to service access and delivery across the world. As a result, individuals increasingly depend on these systems for routine activities such as financial transactions and accessing government services. Even though both sectors benefit from the efficiency and convenience of self-service technologies, their core objectives are different. The private sector mainly emphasizes operational efficiency, profit maximization, and enhanced customer experience through speed and personalization. This results in competition among providers for best designed technology, marketing and diversification. Conversely, the public sector focuses on accessibility, inclusivity, and the effective provision of essential services to all citizens, often within regulated and resource-limited settings. These differing priorities influence the design, implementation, and evaluation of self-service systems in each sector. This paper examines the application of self-service technologies (SST) in public sector digital portals, governmental kiosks, and apps, as well as in private sector ATMs, mobile banking apps, online portals, and hotel check-in apps. A comparative analysis is used to evaluate how these technologies facilitate interactions between customers and technology.

2. Methodology

A qualitative research approach is employed, combining a literature review with structured comparative analysis to examine self-service systems in both private and public sectors. This methodology is designed to facilitate the analysis of system types, sector-specific applications, and the principal differences in technology implementation and user experience.

The research is grounded in a literature review of academic and industry sources concerning self-service technologies, digital service delivery, and user interaction. The reviewed literature offers conceptual definitions and practical examples, facilitating the identification of common system types, including automated teller machines (ATMs), self-checkout systems, online banking platforms, government portals, and ticketing machines. These sources also provide insight into the varied application of such systems based on organizational context and objectives.

The analysis distinguishes between private and public applications of self-service systems. In the private sector, the focus is on efficiency, cost reduction, service speed, and customer experience, with particular attention to technologies such as retail self-checkout, banking applications, and airline kiosks. Conversely, the public sector is examined in terms of accessibility, inclusivity, and service delivery, emphasizing e-government platforms, online public services, and automated service terminals. This approach enables the identification of operational benefits and sector-specific challenges, including over-automation in private services and digital exclusion in public systems.

A comparative analysis evaluates the differences and similarities between the two sectors based on key dimensions, including purpose, user groups, level of innovation, accessibility requirements, and system limitations. Consistent application of these criteria highlights how differing priorities, such as profitability in the private sector and inclusivity in the public sector, influence the design, implementation, and effectiveness of self-service technologies.

Although this methodology offers a structured and comprehensive framework, its primary limitation is reliance on secondary data and existing literature. Nevertheless, it enables a focused and context-sensitive analysis that supports the examination of self-service systems in both sectors.

3. Analysis: Private sector

Banking

In the study of Scherer et al. it is elaborated that the business press frequently commends self-service channels for their significant potential to enhance organizational productivity while simultaneously reducing the costs associated with service delivery. For example, the cost of a banking transaction can be reduced from 1.15 U.S. dollars to 2 cents by transitioning from onsite to online transactions [29]. Additionally, the number of passengers processed for a flight can increase by up to 50 percent through self-check-in options (IATA in SITA 2009). Furthermore, 2.5 employees can be replaced by a single self-checkout kiosk in the grocery store [36]. Forecasts indicate that this trend in business practice will continue, especially in the hospitality and health-care sectors [36] and through the proliferation of mobile self-service applications (L2013). (2015)

Grocery stores

Grocery stores have also been transformed through technological advancements. In large retail environments, the shopping experience can now be fully individualized, eliminating the need for customer-employee interaction. A study of Ittersum et all (2013) explored smart carts as shopping carts equipped with scanners that monitor the total price of a customer's basket during shopping. Consumer surveys indicate that 72% of shoppers would welcome the use of radio-frequency identification tags on products to better track in-store spending (www.Infosys.com). Additionally, 85% of leading retailers identify self-service customer-facing technologies, such as smart shopping carts, as one of the top three opportunities for increasing consumer satisfaction and revenue [33]. Beyond enabling

shoppers to monitor their in-store spending [30], smart shopping carts may enhance customer satisfaction by providing customized and timely promotions, recommending complementary products, sharing nutritional information and recipes, and allowing customers to bypass the checkout lane.

The Kiosk

One widely used self-service technology (SST) is the kiosk [32]. Kiosks have become prevalent in specific segments of the hospitality industry, such as airline self-service check-in and airport information kiosks. Restaurants and hotels have also implemented this technology to serve patrons more efficiently and effectively. In restaurants, self-service kiosks (SSKs) enable customers to place, customize, and pay for food and beverage orders, while hotels use these devices to provide information and facilitate self-check-in and check-out services. Quick service restaurants (QSRs) have particularly embraced SSK technology due to its ability to reduce labor costs [6], improve service speed [21], enhance order accuracy [21], increase sales through up-selling, and appeal to Millennial and younger customers. SSKs allow customers to personalize their meals, exercise greater control over the ordering process, and pay conveniently without waiting in long lines. Approximately 80% of consumers perceive restaurant technology as increasing convenience, while 70% believe it accelerates service and improves order accuracy (NRA, 2017). Perutka (2010) found that customers are willing to accept higher prices and pay an average of $2.47 USD more at QSRs if service is faster (Perutkova, 2010). The promise of shorter waiting times appears to significantly influence customers' decisions to use SSKs [22]. Collectively, these factors indicate that the adoption of SSKs will likely become increasingly significant in the hospitality industry, particularly within the QSR segment, underscoring the relevance of the present study.

Hospitality and travel

Self-service technology has transformed the travel and hospitality sector. Mobile check-in apps streamline the process, reduce waiting times, enhance privacy by minimizing in-person contact, and lower operational costs for hotel chains. However, results of the study of Kattara evidently revealed that hotel customers favor to be served through hotel employees rather than using an innovative technology during guest cycle stages. Results indicated that HIS scenarios were preferred by customers in seven scenarios, representing preregistration, registration, occupancy, and departure stages. In after departure stage, (2013)

Chatbots agents

The Internet has become integral to nearly all aspects of modern life. This development has significantly influenced business purchasing behaviors, contributing to the rapid expansion of electronic commerce (e-commerce). With annual online sales growth rates between 20 and 25%, the e-commerce sector is experiencing substantial acceleration. For consumers, conducting transactions online is associated with speed, efficiency, and a broad selection of products and services, which collectively enhance convenience in daily life. The increasing integration of digital technologies has shifted consumer expectations and needs, necessitating that companies adapt their strategies to the evolving online environment. To address these new market demands, organizations are extending traditional services with digital offerings tailored to customer preferences. Artificial intelligence (AI) plays a central role in this transformation. Applications such as natural language processing, machine learning, robotics, and e-service agents—commonly known as chatbots—are among the most prominent uses of AI in e-commerce. The adoption of chatbots in e-services is particularly notable, as they offer innovative opportunities to enhance customer service. Chatbots function as company

representatives, assisting consumers online by resolving issues, providing information, and offering advice, thereby mitigating dissatisfaction often caused by long wait times in call centers. Ultimately, the objective of deploying chatbots is to fulfill customer needs, which is expected to foster positive attitudes, increase purchase intentions, and promote loyalty, culminating in greater customer satisfaction.

4. Analysis: Public sector

Self-service technology has significantly impacted the public sector by enabling citizens to access government services and information more efficiently. Its adoption has transformed public service delivery, resulting in greater efficiency, cost-effectiveness, and accessibility.

Accessibility

Enhancing accessibility and convenience through self-service technologies has become a central objective within the public sector. Traditional models of public service delivery often rely on in-person interactions or telephone communication, which can impose significant burdens on citizens, particularly those residing in remote areas, individuals with mobility limitations, or those constrained by time. The introduction of self-service technologies, including online portals and automated kiosks, enables users to access essential services independently and at their convenience, regardless of location or time. This shift represents an important step toward greater inclusivity, as it facilitates more equitable access to public services for diverse population groups. Moreover, the convenience associated with self-service technologies contributes to an enhanced citizen experience. In contemporary society, there is a growing expectation for services that are efficient, intuitive, and timesaving. Compared to traditional service delivery methods, digital self-service solutions typically involve fewer procedural steps and significantly reduced waiting times. The implementation of user-friendly interfaces further minimizes confusion and frustration, allowing individuals to navigate complex administrative processes with greater ease. By simplifying service interactions, public sector organizations demonstrate a commitment to improving service quality and respecting citizens' time, which can, in turn, strengthen public satisfaction and trust in governmental institutions. Additionally, the increasing reliance on self-service technologies reflects broader societal shifts toward digitalization, as citizens have become accustomed to seamless digital experiences in other areas of daily life and expect similar standards from public services.

Implementation and Design

According to "Central square" by implementing these technologies, public sector organizations demonstrate their determination to modernize and remain relevant to the needs of a digital society. This is especially critical for engaging younger generations, who are more accustomed to digital interactions and may be hesitant to use old, paper-based systems. Improving accessibility and convenience with self-service technologies requires more than just staying current with technological improvements. It is about radically changing the way services are offered so that they are more inclusive, efficient, and meet the demands of today's society. It signifies a shift towards a more citizen-centered approach, in which the public's wants and preferences are prioritized in service design and delivery.

A well-designed, user-oriented approach not only improves efficiency but also minimizes the confusion and frustration often associated with complex administrative procedures. By streamlining these processes, public sector organizations demonstrate a clear commitment to valuing citizens' time and effort, which can contribute to higher levels of

satisfaction and increased trust in government services. Furthermore, the growing adoption of self-service technologies reflects changing public expectations. As digital solutions have become standard across many areas of everyday life, citizens increasingly anticipate similar levels of efficiency and accessibility in public services. The integration of such technologies signals a willingness on the part of public institutions to modernize and adapt to the demands of a digitally engaged population. This transformation is particularly important for engaging younger generations, who are more accustomed to digital interactions and may be less inclined to use traditional, paper-based systems. However, the shift toward self-service technologies extends beyond simply adopting new tools; it represents a broader reconfiguration of service delivery models. Ultimately, this transition reflects a move toward a more citizen-centered approach, in which services are designed and delivered with a primary focus on the needs, preferences, and expectations of the public, ensuring greater inclusivity, efficiency, and responsiveness in contemporary society.

Adoption and convenience

Recent evidence highlights several persistent challenges in the adoption of digital public services. A study conducted by Deloitte, based on a survey of 1,000 residents in the United States, found that only a small proportion of citizens—approximately 23 percent—regularly engage with local government services through digital platforms. This limited level of interaction reflects broader concerns related to usability and service quality. Respondents identified issues such as complex website navigation, comparatively less advanced digital services than those offered in the private sector, and ongoing barriers to accessibility.

One of the most significant challenges lies in ensuring equal access to these technologies. Variations in digital literacy, as well as disparities in access to reliable internet connections and appropriate devices, continue to limit the effectiveness of self-service systems in the public sector. In response, governments have introduced alternative access points, including public libraries and community centers, where citizens can use digital services. Support initiatives, such as training programs and user assistance, are also employed to improve confidence and competence among users.

At the same time, the implementation of digital service systems presents notable financial and operational demands. Beyond the initial investment in technological infrastructure, ongoing costs related to maintenance, updates, and system improvements must also be considered. To address these constraints, public institutions often explore collaborative arrangements with private sector partners or seek funding through dedicated programs that support digital transformation. Gradual implementation strategies are also commonly adopted, allowing organizations to manage expenditures more effectively while incorporating feedback from early stages of deployment.

In addition, issues of data protection and privacy remain central to the successful operation of digital public services. The handling of sensitive personal information requires strict adherence to legal and regulatory standards, as well as continuous investment in secure technological systems. Ensuring transparency in how data is collected, stored, and used is essential for maintaining public confidence and fostering trust in government-led digital initiatives.

5. Discussion

The findings of this study reveal a distinct divergence in the development and utilization of self-service systems between the private and public sectors, despite their common technological basis. In the private sector, the adoption of self-service technologies is primarily motivated by goals of efficiency, cost reduction, and competitive advantage. Examples such

as online banking, self-checkout systems, and hospitality kiosks illustrate how organizations invest in these technologies to streamline operations, minimize labor requirements, and improve customer experience. Innovation in this context is rapid and shaped by market competition, leading to highly optimized, user-focused systems that emphasize speed, personalization, and convenience.

In contrast, the public sector implements self-service technologies with the primary objective of ensuring accessibility and equitable service delivery. Although digital platforms such as e-government portals and automated service systems have improved efficiency, their deployment is constrained by broader social responsibilities. Challenges such as digital literacy, unequal access to technology, and the necessity for inclusive design persist. Unlike the private sector, where adoption is typically optional and driven by customer preferences, public sector systems are required to serve all citizens, including individuals with limited technological proficiency.

A significant distinction exists in the pace of innovation and system development between sectors. Private organizations implement and refine technologies rapidly in response to consumer demand. In contrast, public institutions frequently operate within regulatory, financial, and bureaucratic constraints that impede digital transformation. This disparity contributes to user dissatisfaction, as evidence suggests that public digital services are often perceived as less advanced and less intuitive than those offered by the private sector.

Both sectors encounter similar challenges, particularly regarding data security, system reliability, and user trust. However, the consequences of failure differ substantially. In the private sector, service disruptions may lead to financial losses or diminished customer loyalty. In the public sector, such failures can erode trust in institutions and restrict access to essential services.

In summary, while self-service technologies provide substantial benefits in both sectors, their effectiveness is influenced by sector-specific priorities. The private sector demonstrates strengths in innovation and efficiency. Conversely, the public sector is essential for promoting inclusivity and equitable access. These findings underscore the importance of adopting a balanced and integrated approach to technology implementation.

6. Conclusion

Self-service systems are now widely used in both private and public sectors, reflecting a broader shift to digital services. These technologies offer benefits such as higher efficiency, lower costs, and greater convenience, but their impact differs by sector. Private companies focus on innovation, speed, and improving customer experience, while public organizations prioritize making services accessible and inclusive.

Still, there are ongoing challenges like digital inequality, usability issues, and worries about trust and security. These problems show that technology alone is not enough. Successful implementation also needs to address user needs and wider social issues. In the end, self-service systems work best when they balance efficiency with accessibility. Bringing together private sector innovation and public sector inclusivity is the most effective way to develop future services development.

References

1. Aarlott, Espen. n.d. The User Experience of Self-Service Systems. Norwegian University of Science and Technology.
2. Aloudat, A., Michael J. Michael, and Katina Michael. 2017. Privacy and Security in Multi-User Health Kiosks.

3. Anderson, Ross. 2020. Security Engineering: A Guide to Building Dependable Distributed Systems. 3rd ed. Hoboken, NJ: Wiley.
4. Ba, Sulin, Andrew B. Whinston, and Han Zhang. 2010. "Building Trust in Online Services." Information Systems Research.
5. Barnes, James G. 1997. "Closeness, Strength, and Satisfaction: Examining the Nature of Relationships between Providers of Financial Services and Their Retail Customers." Psychology & Marketing.
6. Beatson, Amanda, Nicole Lee, and Leonard V. Coote. 2007. "Self-Service Technology and the Service Encounter." Service Industries Journal 27 (1): 75–89.
7. Bitner, Mary Jo, Amy L. Ostrom, and Matthew L. Meuter. 2000. "Implementing Successful Self-Service Technologies." Academy of Management Executive.
8. Campbell, Dennis, and Francesca Frei. 2010. "Cost Structure, Customer Profitability, and Retention Implications of Self-Service Distribution Channels." Management Science.
9. Collier, Joel E., and Sheryl E. Kimes. 2013. "Only If It Is Convenient: Understanding How Convenience Influences Self-Service Technology Evaluation." Journal of Service Research.
10. Dabholkar, Pratibha A. 1996. "Consumer Evaluations of New Technology-Based Self-Service Options." International Journal of Research in Marketing.
11. Deloitte. 2024. Digital Citizen Survey: U.S. Report. https://www.deloitte.com/content/dam/insights/articles/2024/in176014_cgi_digital-citizen-survey/DI_Digital-citizen-us.pdf
12. Ennew, Christine T., and Martin R. Binks. 1999. "Impact of Participative Service Relationships on Quality, Satisfaction and Retention." Journal of Business Research.
13. European Union. 2016. General Data Protection Regulation (GDPR).
14. Fitzsimmons, James A. 1985. "Consumer Participation and Productivity in Service Operations." Interfaces.
15. Hitt, Lorin M., and Frances X. Frei. 2002. "Do Better Customers Utilize Electronic Distribution Channels?" Management Science.
16. Infosys. 2007. Retail Technology Consumer Survey.
17. Ibitola, Joseph. 2023. Real-Time Transaction Security and Payment Vulnerabilities. Flagright.
18. Ittersum, Koert van, et al. 2013. "Smart Shopping Carts: Consumer Behavior and Retail Technology."
19. Jain, Anil K., Arun Ross, and Salil Prabhakar. 2016. A Framework for Biometric ATM Authentication.
20. Kattara, Hanan Saad, and Osman Ahmed El-Said. 2013. "Customers' Preferences for New Technology-Based Self-Services versus Human Interaction Services in Hotels." Tourism and Hospitality Research 13 (2): 67–82.
21. Kincaid, C. Scott, and Seyhmus Baloglu. 2005. "An Exploratory Study on the Impact of Self-Service Technology on Restaurant Operations." Journal of Foodservice Business Research 8 (3): 55–65.
22. Kokkinou, Aikaterini, and David A. Cranage. 2015. "Why Wait? The Effect of Waiting Time on Customer Behavior in Self-Service Technology." International Journal of Hospitality Management 47: 20–28.
23. Komaragiri, Venkata Bhardwaj, and Andrew Edward. 2022. "AI-Driven Vulnerability Management and Automated Threat Mitigation." International Journal of Scientific Research and Management 10 (10): EC-2022-980–998.
24. Kumar, R., and Ritu. 2020. "Customers' Perception towards Security and Privacy Issues of ATMs." International Journal for Research in Engineering Application & Management.
25. Lovelock, Christopher H., and Robert F. Young. 1979. "Look to Consumers to Increase Productivity." Harvard Business Review.
26. Meuter, Matthew L., Amy L. Ostrom, Robert I. Roundtree, and Mary Jo Bitner. 2000. "Self-Service Technologies: Understanding Customer Satisfaction with Technology-Based Service Encounters." Journal of Marketing.
27. Meuter, Matthew L., et al. 2005. "Choosing among Alternative Service Delivery Modes." Journal of Marketing.
28. Mills, Peter K., Richard B. Chase, and Nancy Margulies. 1983. "Motivating the Client/Employee System." Academy of Management Review.

29. Moon, Youngme, and Frances X. Frei. 2000. "Exploding the Self-Service Myth." Harvard Business Review.
30. Nelson, Paul. 2008. "Retail Self-Service Technology Trends."
31. Perutka, David. 2010. "The Value of Speed in Quick Service Restaurants." Hospitality Technology Magazine.
32. Rowley, Jennifer, and Frances Slack. 2003. "Kiosks 2.0: The Role of Self-Service Technology in Service Delivery." International Journal of Retail & Distribution Management 31 (2): 95–104.
33. Rosenblum, Paula. 2007. "Retail Technology Adoption Trends."
34. Scherer, Anne, Nancy V. Wünderlich, and Florian von Wangenheim. 2015. "The Value of Self-Service: Long-Term Effects of Technology-Based Self-Service Usage on Customer Retention." MIS Quarterly 39 (1): 177–200. https://www.jstor.org/stable/26628346
35. Simon, Franck, and Jean-Claude Usunier. 2007. "Cognitive, Demographic, and Situational Determinants of Service Customer Preference for Personnel-in-Contact over Self-Service Technology." International Journal of Research in Marketing.
36. The Economist. 2009. "Automation and Self-Service Trends in Retail and Hospitality."
37. UpGuard. 2024. The Biggest Data Breaches in Financial Services.
38. Venkateswaran, Aishwarya. 2020. Impact of Self-Service Technology on Passenger Satisfaction.
39. Wang, Rong, Lloyd C. Harris, and Paul G. Patterson. 2012. "Customer Choice of Self-Service Technology: The Role of Situational Influences and Past Experience." Journal of Service Management 23 (1): 54–78.
40. Xue, Ming, Chris Forman, and others. 2007. "Customer Adoption of Self-Service Technologies."
41. Zhu, Feng, et al. 2007. "Information Transparency and Customer Behavior.

Security and Privacy Concerns in Self-Service Systems

Petar Genov

University of National and World Economy, Sofia, Bulgaria
petar.genov@unwe.bg

Abstract. This paper explores the security and privacy concerns in self-service systems (SSSs). Within the constant implementation of such technologies constant call for action is required from companies to provide, protect and enable safe use. The study focuses on weak authentication, physical attacks on devices and network and software vulnerabilities. The frame of this study is designed to dive deeper onto the existing problems and discuss ways of handling such errors.

Keywords. Self-service systems, privacy, security, public sector, private sector.

1. Introduction

Contemporary society is experiencing a significant shift toward technology-enabled service access, as organizations and companies increasingly implement self-service systems (SSSs). Advances in information technology have facilitated the development of diverse service channels that empower users to complete tasks independently. Self-service systems are now integral to daily life, enabling transactions without direct human assistance. Individuals frequently interact with these technologies, often without conscious awareness, for activities such as online bill payment, self-checkout at retail stores, or automated fuel dispensing [1].

The widespread adoption of SSSs across industries is primarily driven by improved efficiency, reduced operational costs, and enhanced user convenience. These systems offer organizations a competitive advantage by streamlining service delivery and reducing reliance on human resources.

However, alongside these benefits, the increasing reliance on self-service technologies raises critical concerns regarding security and privacy. As users are required to input sensitive personal and financial information, these systems become potential targets for cyberattacks, fraud, and data misuse. Furthermore, the lack of direct human oversight and varying levels of user awareness can exacerbate existing vulnerabilities.

2. Methodology

This study uses a qualitative research approach based on a critical review of academic literature and case-based analysis. The goal is to identify common security and privacy concerns in self-service systems (SSSs) and to see how these issues are presented in different technologies and user settings.

The main method is a systematic literature review. The studies chosen cover areas like human-computer interaction, cybersecurity, and service system design. For example, Aarlott's work on user experience in self-service systems shows how perceived security and privacy risks affect user trust and adoption [1]. Kumar and Ritu (2020) also look at how people view ATM security, showing that trust in self-service systems is linked to how vulnerable users feel and their past experiences with risk.

To further add on these perspectives, the study incorporates research focused on technical and infrastructural vulnerabilities. Aloudat et al. [2] identify critical weaknesses in multi-user kiosk environments, particularly in relation to shared access points and insufficient data protection mechanisms. In parallel, Jain et al. [6] explore biometric authentication frameworks in ATM systems, emphasizing both their security advantages and the privacy risks associated with storing sensitive biometric data. These studies collectively illustrate that technological solutions often introduce new layers of complexity rather than fully resolving existing risks.

An analytical approach is applied to synthesize findings across the literature. Key themes—including authentication, data protection, system design, and human factors—were not predefined but emerged through repeated patterns in the reviewed studies. For example, Neisse et al. highlight the increasing security challenges in interconnected systems, suggesting that the expansion of networked self-service technologies amplifies exposure to cyber threats. At the same time, Venkateswaran [10] demonstrates that user satisfaction in self-service environments is strongly influenced by perceived security, reinforcing the idea that technical robustness alone is insufficient without user trust.

Crucially, this approach takes a critical stance on the literature. Fewer studies focus on the larger socio-technical dimension, which includes user behavior, awareness, and ethical consequences, whereas many highlight technological solutions like encryption or biometrics . This disparity points to a research gap where human aspects are frequently overlooked despite being a major source of vulnerability.

But there are a few restrictions to be aware of. The study is limited to issues that have already been identified and may miss new or context-specific dangers due to the dependence on secondary data. Furthermore, direct comparison is difficult due to differences in research scope and methods across the chosen sources. Validating results in real-world settings is also hampered by the lack of primary empirical data, such as user surveys or system testing.

Despite these limitations, the chosen methodology provides a comprehensive and critically informed framework for analyzing security and privacy concerns in self-service systems. By integrating technical, user-centered, and contextual perspectives, it enables a more nuanced understanding of the risks and challenges associated with these increasingly prevalent technologies.

3. Analysis of Security and Privacy Risks

In this section the paper will focus into key risks surrounding the implementation of SST technologies. This paper identifies SST technologies as described by Aloudat et al. [2] as technological interfaces that enable clients to complete tasks independently of service staff. In a broader sense, self-service refers to any technology-based exchange or transaction between a client and a business in which the client is the sole human involved.

Unauthorized Access / Weak Authentication

The widespread adoption of ATMs has encouraged the banking sector to develop alternative delivery channels, such as internet banking, mobile banking, green channel counters, kiosk banking, credit cards, debit cards, and online bill payment services. Across these channels, Upguard claims that users may encounter situations in which unauthorized individuals gain access to their income or savings. Security failures in financial systems can result in large-scale data breaches with severe consequences for both organizations and users. For instance, a major incident in the financial sector involved a breach affecting over 100 million individuals, exposing sensitive information including social security numbers, bank account details, and personal identification data. Such breaches frequently result from

misconfigured security systems, such as inadequately secured cloud storage and firewall vulnerabilities, which provide entry points for attackers. These incidents illustrate that even large, well-established institutions may fail to adequately protect user data, emphasizing the necessity of robust system configuration and continuous security monitoring. Once exposed, sensitive information can be exploited for identity theft, financial fraud, and long-term privacy violations, significantly eroding user trust in self-service systems [9].

Even in the absence of major security incidents, proactively strengthening account security is more effective than relying on the hope that attacks will not succeed. Evidence indicates that users who implement secondary verification steps, such as multi-factor authentication (MFA), and utilize strong, unique passwords, as recommended by iOS, can significantly reduce vulnerabilities across self-service technologies (SST).

Physical Attacks (ATMs, kiosks, skimming, tampering)

According to the official website of FBI It is estimated that skimming costs financial institutions and consumers more than $1 billion each year. Skimming occurs when devices are unlawfully installed on or within ATMs, point-of-sale (POS) terminals or petrol pumps, capturing card data and recording cardholder Pins. Criminals exploit the information to generate bogus payment cards and then make unauthorized purchases or steal from victims' accounts.

ATM and point-of-sale (POS) skimming attacks are a key category of security concerns in self-service systems, as they highlight how physical and technical weaknesses can be exploited concurrently. Skimming devices can be quietly inserted within or over card readers to acquire magnetic stripe data, while additional tools like pinhole cameras or keypad overlays are utilized to record users' PIN entries. These assaults take little time to carry out and can occur in ordinary settings like retail establishments or freestanding ATMs. More modern skimming devices can communicate stolen data wirelessly in real time, enhancing the efficiency and scope of fraud.

Incidents like these underscore the continuing risks of legacy technologies, especially magnetic stripe cards, which remain more vulnerable than chip-enabled solutions. Vulnerable populations are disproportionately affected by security flaws, especially when limited fraud protections make it difficult or impossible to get paid back, as seen in targeted attacks on systems like electronic benefits transfer (EBT) cards. These results build on a wider range of work which suggests that the security of self-service systems is determined by more than just technical safeguards, and also includes physical design, user awareness and the systems in place to respond.

Malware & Software Vulnerabilities

Malware and software vulnerabilities pose significant challenges to the security of self-service technologies that rely on complex, interconnected network infrastructures [7]. Komaragiri and Edward note that “traditional security approaches, which depend on a fragmented combination of tools such as firewalls, network monitoring systems, and static access control rules, commonly fail to provide extensive protection due to their lack of integration and adaptability. This fragmented architecture not only increases the complexity of security management but also causes new vulnerabilities, especially when systems trust in old software or poorly maintained security configurations” [7]. In SST ecosystems such as online platforms, attackers typically take advantage of such system weaknesses and deploy malware to steal vital user data. Thus, with the mobilization of all services and more human presence online the risks of such malware attacks increase daily. Recent studies highlight the potential of artificial intelligence (AI) and machine learning to enhance cybersecurity through

real-time anomaly detection and automated threat mitigation. Machine learning can process large volumes of network data, identify unusual patterns, and respond to threats more efficiently than traditional rule-based methods. However, the adoption of these advanced security solutions is often limited by high implementation costs and operational complexity, particularly in large-scale systems [7]. While AI offers significant benefits for securing self-service systems, there remains a need for more integrated, scalable, and adaptive approaches to managing software vulnerabilities in digital service environments.

Network Vulnerabilities and Data Interception

Real-time transactions power modern digital payments but expose businesses to critical security vulnerabilities including interception attacks, man-in-the-middle exploits, phishing schemes, and identity theft. In an article by Flagright 2023, the author mentions that real-time transaction systems, which underpin modern self-service technologies such as mobile banking, digital wallets, and online payment platforms, introduce a distinct set of software and network vulnerabilities due to their speed and complexity. The instantaneous nature of these transactions significantly reduces the time available for fraud detection, creating opportunities for attacks such as man-in-the-middle exploits, data interception, phishing, and identity theft. These systems rely on multi-step processes involving authentication, data transmission, and validation across multiple stakeholders, each representing a potential point of failure. In particular, vulnerabilities in payment gateways—such as insecure APIs, weak encryption during data transmission, and session hijacking—can allow attackers to access or manipulate sensitive financial information. Furthermore, the reliance on interconnected digital infrastructures increases exposure to large-scale attacks, including distributed denial-of-service (DDoS) incidents that can disrupt services and mask fraudulent activities. To address these risks, contemporary research emphasizes the growing role of artificial intelligence in transaction security, where machine learning models analyze behavioral patterns and detect anomalies in real time. While these AI-driven approaches enhance fraud detection capabilities, they also introduce new challenges related to implementation complexity and data dependency, reinforcing the need for a multi-layered and adaptive security framework in self-service environments [5].

4. Discussion and Conclusion

The widespread adoption of self-service systems has significantly transformed service interactions by providing efficiency, speed, and convenience across various sectors. Nevertheless, as explored in this paper, these benefits are surrounded by a complex array of security and privacy challenges that call for careful and further consideration. Vulnerabilities in self-service technologies arise from the interaction of technical weaknesses, system design limitations, and human factors.

Security threats, including unauthorized access, physical attacks, malware exploitation, and network-based vulnerabilities, underscore the ongoing exposure of self-service systems despite technological progress. Although advancements, the continued presence of weak authentication mechanisms, outdated software, and insecure data transmission points to systemic issues in system development and maintenance. Although technological solutions such as biometric authentication and AI-driven threat detection provide notable advancements, they do not fully mitigate risk and may add additional complexity that necessitates careful implementation and sustained oversight.

A central finding of this study is that security in self-service environments cannot be achieved solely through technical interventions. User behavior, awareness, and trust are equally critical in shaping overall system resilience. Users' preference for convenience,

coupled with limited risk awareness, introduces vulnerabilities that technical measures alone cannot address. Furthermore, organizations must balance usability and security, as excessive complexity may deter adoption or encourage unsafe practices.

Moreover, the increasing dependence on real-time digital transactions and interconnected platforms indicates that the scale and impact of potential security breaches are likely to escalate. As shown in the analysis, even established institutions remain vulnerable to failure, with consequences that extend beyond financial loss to encompass long-term reputational harm and diminished user trust.

These findings underscore the necessity of a holistic approach to security. Effective protection of self-service systems requires the integration of robust technological measures, user-centered design, continuous monitoring, and well-defined regulatory frameworks. Future advancements should prioritize not only the enhancement of technical defenses but also the mitigation of human and organizational factors contributing to system vulnerability.

Ultimately, the success of self-service technologies relies not only on efficiency but also on the provision of secure and trustworthy environments for users. Without this foundation, the long-term sustainability and acceptance of these systems remain uncertain.

References

1. Aarlott, Espen. n.d. The User Experience of Self-Service Systems. Norwegian University of Science and Technology. https://www.ntnu.edu/documents/139799/1270604448/Artikkel%2BEspenAarlott.pdf
2. Aloudat, A., Michael J. Michael, and Katina Michael. 2017. Privacy and Security in Multi-User Health Kiosks. https://www.researchgate.net/publication/318715645
3. Centier Bank. 2023. How Safe Are Mobile Banking Apps—Really? https://www.centier.com/resources/articles/article-details/how-safe-are-mobile-banking-apps-really
4. Federal Bureau of Investigation (FBI). 2024. ATM and POS Skimming. https://www.fbi.gov/how-we-can-help-you/scams-and-safety/common-frauds-and-scams/skimming
5. Ibitola, Joseph. 2023. Real-Time Transaction Security and Payment Vulnerabilities. Flagright.
6. Jain, Anil K., Arun Ross, and Salil Prabhakar. 2016. A Framework for Biometric ATM Authentication. https://arxiv.org/abs/1602.04620
7. Komaragiri, Venkata Bhardwaj, and Andrew Edward. 2022. AI-Driven Vulnerability Management and Automated Threat Mitigation. International Journal of Scientific Research and Management 10 (10): EC-2022-980–998. https://doi.org/10.18535/ijsrm/v10i10.ec05
8. Kumar, R., and Ritu. 2020. Customers' Perception towards Security and Privacy Issues of ATMs. International Journal for Research in Engineering Application & Management. https://ijream.org/papers/IJREAMV06I0161058.pdf
9. UpGuard. 2024. The Biggest Data Breaches in Financial Services. https://www.upguard.com/blog/biggest-data-breaches-financial-services
10. Venkateswaran, Aishwarya. 2020. Impact of Self-Service Technology on Passenger Satisfaction. https://norma.ncirl.ie/4690

Opportunities for Increasing the Potential of the Bulgarian Defense Intelligence Service through the Voluntary Active Reserve of the Bulgarian Armed Forces

Stanimir Stoyanov

University of National and World Economy, Sofia, Bulgaria
Stanimir.stoyanov@unwe.bg

Abstract. The security environment around the world and in Bulgaria requires an unconventional approach in providing additional expertise to the employees of the special services. The upcoming changes in the regulatory framework for the appointment of personnel in the "Defense Intelligence Service" (SVR) require the full use of the opportunities offered by the voluntary reserve of the armed forces. Using this special approach, the service can only increase its potential and deploy a wider range of capabilities, positively assessed by the state leadership and international partners.

Keywords. Military intelligence, voluntary reserve, capabilities, amendments to the regulations, armed forces.

1. Introduction

Modern economic development and the transformation of the modern world hide a number of risks and threats to the achievement of our strategic national goals. The dynamic geopolitical situation, the globalization of international relations and the growing interconnectedness of the private economic and defense sectors require a proactive approach to increasing the training of personnel in the field of security and, more specifically, in the law enforcement agencies. Many countries and their law enforcement institutions resort to innovative methods to expand the knowledge and complement the spectrum of skills of current and future employees of the special services, paying attention to the one dealing with military intelligence functions.

A less frequently used method is presented, which can undoubtedly strengthen the capabilities of the personnel and provide new opportunities for the implementation of the main and additional activities. For this reason, proposals for amendments to the regulations on the organization and activities of the Military Intelligence Service (DIS/SVR) should be selected, providing more options for binding citizens to the structure. The main aspect that is being addressed is the possibility of strengthening the unit's capabilities through the voluntary reserve of the armed forces, where the SVR has its own independent staff.

2. Current status of the regulatory framework and organizational structure

The Military Intelligence Service is a specialized structure within the Ministry of Defense, directly subordinate to the Minister of Defense. Due to its specificity, the service is regulated by a special law, as well as regulated for its implementation. This is one of the main units dealing with national security and, accordingly, a large part of the information regarding its structure, organization, number, staff, operational activities, personnel and level of training

is classified. According to the Military Intelligence Law (1) the institution unites all intelligence formations in the Armed Forces of the Republic of Bulgaria. It functions with the aim of obtaining, processing and providing intelligence information in the interests of the country's defense and national security. Military intelligence participates in the activities of the intelligence communities of the North Atlantic Treaty Organization (NATO) and the European Union (EU), carries out intelligence exchange and intelligence support in allied and coalition operations. It is important to note that the main activity includes, in addition to the organization, the intelligence military formations in the Bulgarian Army, as well as intelligence units in the governing bodies. As a politically neutral, objective and impartial institution, military intelligence is guided by principles laid down by the Constitution, the laws of the country, the international treaties to which Bulgaria has undertaken a commitment, respect for the freedoms of citizens, interaction with state authorities and public order services, and centralized management of intelligence activities. As part of the system for protecting national security, the institution combines overt and covert forms of work.

Due to the specifics of its activities, the service is an independent legal entity, and its director is a secondary budget manager under the Minister of Defense. The service processes and analyzes all information provided by the intelligence structures of the Bulgarian Army, NATO and the EU, by the intelligence agencies of the NATO and EU member states, and by other partner services, and is autonomous in ensuring the protection of classified information, internal security and counterintelligence support for operational intelligence activities. Military intelligence cannot be assigned tasks of an internal political nature. It does not collect, process, analyze or store information of an internal political nature. The main tasks of the service at the strategic level are:

- obtaining and analyzing strategic intelligence information necessary for the defense of the country and the protection of national security from external encroachments, risks and threats;
- providing information for the activities of state institutions in the field of security and defense of the country;
- carry out continuous monitoring, early detection and warning of risks and threats to national security and collective defense; (7)
- providing intelligence information for planning needs in the armed forces and for the development of operational capabilities;
- developing information and operational cooperation with the intelligence structures of NATO and the EU, with the intelligence bodies of the NATO and EU member states, and with partner services in the interest of national and collective defense and security;
- implementing the exchange of intelligence information and conducting joint operations with the intelligence structures of NATO and the EU, the intelligence bodies of the NATO and EU member states, and with partner services;
- providing intelligence information in the interest of the security of operations and missions outside the territory of the country, in which Bulgarian military contingents and individual servicemen participate;
- implementing direct management of the service of the military attaché to the foreign missions of the Republic of Bulgaria.

To fulfill its assigned tasks, the service:

* uses specific forms, methods and means of operational work
* attracts Bulgarian citizens or citizens of another country with their voluntary consent to cooperate in performing intelligence tasks
* provides measures for the protection and security of persons who work or have worked, cooperate or have cooperated with military intelligence

* creates its own coverts for its employees

* uses coverts in departments, organizations, commercial companies and other legal entities

* draws up and uses documents for foreign identity and other official documents necessary for the covert of employees of the service who perform operational intelligence activities, including reflecting the necessary changes in electronic databases

* develops and produces special equipment, auxiliary means and documents to ensure its operational activities

* uses documents to protect and conceal the belonging of its personnel, buildings and facilities

* participates in the organization and implementation of military-diplomatic activity

* deploys and uses electronic intelligence bodies to obtain information in the interest of national security from external electronic systems and means

* develops capabilities to obtain intelligence information, independently or jointly with NATO, EU and partner services, using space, aviation, sea or land systems for obtaining images

* concludes contracts with legal entities and individuals for services and supplies related to its activities, in accordance with the Public Procurement Act and the Classified Information Protection Act.

Although the service is organized into management, directorates, departments and sectors, the Minister of Defense exercises the direct management of military intelligence at the strategic level and the general management of military intelligence in the Bulgarian Army. In addition, he sets information tasks; issues regulatory acts related to the activities of military intelligence; exercises control over the financial and administrative activities and the disposal of the material and technical resources of the service; and submits to the Council of Ministers proposals for the adoption of regulatory acts related to the unit and an annual report on the implementation of the assigned tasks (6).

The employees in military intelligence are divided into three categories: military personnel; civilian employees under official legal relationship in accordance with the Civil Servant Act and civilian employees under employment legal relationship. The status of the employees is regulated in accordance with the Law on Defense and Armed Forces of the Republic of Bulgaria, the Regulations for the Implementation of the Law on Defense and Armed Forces of the Republic of Bulgaria, the Law on Civil Servant, the Labor Code, the Law on the Defense and Armed Forces of the Republic of Bulgaria and the Regulations on the Organization and Activities of the SVR. All except military personnel perform functions related to ensuring the activities of the institution, which are different from the tasks of military intelligence under Art. 12 of the Law on the Defense and Armed Forces of the Republic of Bulgaria. If necessary, the director of the service may also determine positions in which civilian employees perform tasks under Art. 12 of the Law on the Internal Security Service. To fulfill their official duties, personnel may hold positions undercover.

In case more in-depth attention is paid to the internal order, some more specific details stand out. Regarding the official time, breaks and leaves of military personnel are determined under the terms and conditions of the Law on the Defense and Armed Forces of the Republic of Bulgaria and the regulations on the organization and activities of the Internal Security Service. The working time, breaks and leaves of civilian employees - under the terms and conditions of the Law on the Civil Servant (9), respectively the Labor Code and the regulations on the organization and activities of the Internal Security Service. The director of the institution may determine a specific organization of the official time of military personnel and the working time of civilian employees whose permanent workplaces are located in military

formations or facilities located in high-altitude areas outside populated areas, in compliance with the general rules for ensuring healthy and safe working conditions. The financial relations between the service and the employees who hold undercover positions are regulated under the conditions and in accordance with the procedure determined by the regulations on the organization and activities of the SVR(2). Regarding the additional requirements, it is clear that after being released from military service or terminating the legal relationship, military personnel or civilian employees of the organization do not have the right, without the permission of the director or an official authorized by him, to seek or establish contacts with persons who became known to them during or in connection with the performance of operational intelligence activities. Military personnel and civilian employees, upon being released from military service or terminating their legal relationship, sign a declaration under para. 1 of the SVR Act (Military Intelligence Law). The employees of the unit are obliged to protect state and official secrets even after their release from military service or terminating their legal relationship in accordance with the Classified Information Protection Act.

The requirement to use a polyphysiograph on former and current personnel at the discretion of the organization's management should not be ignored. The structure is staffed independently under the conditions and in accordance with the procedure specified in the regulations for the structure and activities of the service. An employee working in a body of the national security protection system may be accepted for military service or appointed to a position, if he meets the requirements and conditions for holding it and after concluding a written agreement between him, the director of the service and the head of the relevant body, and is not liable for reimbursement of the costs of maintenance, training, qualification and/or retraining. The requirement is that persons who meet the Law on Defense and the Armed Forces of the Republic of Bulgaria (8) are accepted for military service; but also specific requirements determined by an act of the Minister of Defense. Acceptance for military service and appointment to a position in the organization are carried out after selection, including through psychological testing. The test is carried out to establish psychological compatibility with the requirements of the relevant field of activity. Only candidates who are psychologically compatible with the requirements of the field of activity are appointed. On the other hand, only persons who have received permission from the information security officer to access the level of classified information required for the position are accepted for military service. To perform intelligence tasks, the Director of the SVR may appoint persons who are in employment or service relationships with another employer or appointing authority, as well as dismiss them from office and from military service. The requirements, conditions and procedure for acceptance for military service are determined by an act of the Minister of Defense. The training of personnel for the needs of the institution is carried out in higher military or civilian schools in the country and abroad. In addition, specialized centers of the Ministry of Defense may also be used. If necessary, for international training, individual regulation by the Minister of the principal department is required (2).

3. Proposals for changes in regulatory documents and opportunities to strengthen staff potential

The accelerated movement of dynamics in the security environment also requires a flexible approach to providing personnel for the Military Intelligence Service. Bearing this in mind, the requirements for personnel in relation to all the regulatory documents listed above may be reduced.

With the appointment of a civilian employee, the institution can provide it with project employment/short-term contracts, in which the conditions are previously agreed with all

parties involved. If we are "talking" about contracts for an official legal relationship, an additional regulation can be inserted through the Ministry of Defense, valid only for certain positions in the service, where special conditions may be present in the presence of specific circumstances. This regulation could be valid for both the special and the overseas staff in the unit. Adding a change in which control of these can be exercised by a deputy director in the service, significantly reduces the time for making a decision and the bureaucratic obstacles in the chain of command. In the presence of civilian employees or military personnel appointed on a project basis, framework contracts could be fixed, but according to the requirements, circumstances and conditions of appointment, an additional annex must also be present, which should satisfy the two/three parties with clearly expressed parameters. The last part is emphasized, because if respected experts are sought, these would be people with positions, knowledge, opportunities and influence in their field of activity. If such people are disappointed, this will inevitably lead to a negative attitude towards the structure, which will spread exponentially in the security circles. It is usually important whether such people have the opportunity to share even the name of the structure where they perform tasks, sometimes maybe temporary ones. In one of the magazines dedicated to national security, other reasons are noted that would encourage many external experts to cooperate with the SVR, provided that they feel enough freedom, can present their work in a more free form and are not subject to too many bureaucratic restrictions (3).

Another positive outcome could be the increasing of the resilience by strengthening the capacity of the critical subsystems with more frequent drills and testing plans (5).

If the reliability check by the information security team is successfully completed, a section could be implemented where data is obtained from associates with a lower level of clearance, but working in various directions. This would be especially appropriate for reservists temporarily appointed to the institution. Speaking of voluntary reserve, the key word here is "flexibility". All reservist contracts should have flexible terms, not only for the sake of the candidate's satisfaction, but also due to the specifics of the service. Provided that an agreement is reached in which reservists collect information, especially outside the country, each cadre must have sufficient confidence that he is legally and physically protected by the service. The annexes to the framework contracts or reservist agreements must correctly describe both the risks and the benefits for all parties. Reservists would require both to be well paid for their efforts and to be satisfied, through recognition, but also freedom to then carry out their main activity without restrictions. For example, the case where a representative of an academic organization is stopped by a native power structure from contacting services and embassies of other countries, just because he was carrying out activities in the Bulgarian interest, happened not that long ago.

Excluding the situation of vertical dependence, and rather cooperation with the military-intelligence structure horizontally (even through a reservist contract), a much more positive response is expected from each candidate. Selecting from a list the directions in which potential employees are available, one can also include:

- University professors, traveling frequently and having a connection to the security field. This type of people would rather want to be in partnership with the service, rather than depending on it, due to their accumulated positions in society. Their motivation, in addition to gaining additional recognition through their projects together with the SVR, would be mostly financial. They can help both with expertise, especially in various aspects: nuclear, chemical, biological and other technologies, and also with influence in foreign organizations with which they work together, in the form of institutes, accreditation agencies, research universities, state institutions, etc.

- University professors, focused on students, having a connection to security. These guild representatives could easily screen out suitable collaborators for the service by observing them in the educational environment, due to their less frequent engagement with other projects. The motivation for working with the SVR would be both reputational (to be more recognizable and validated in society) and financial.

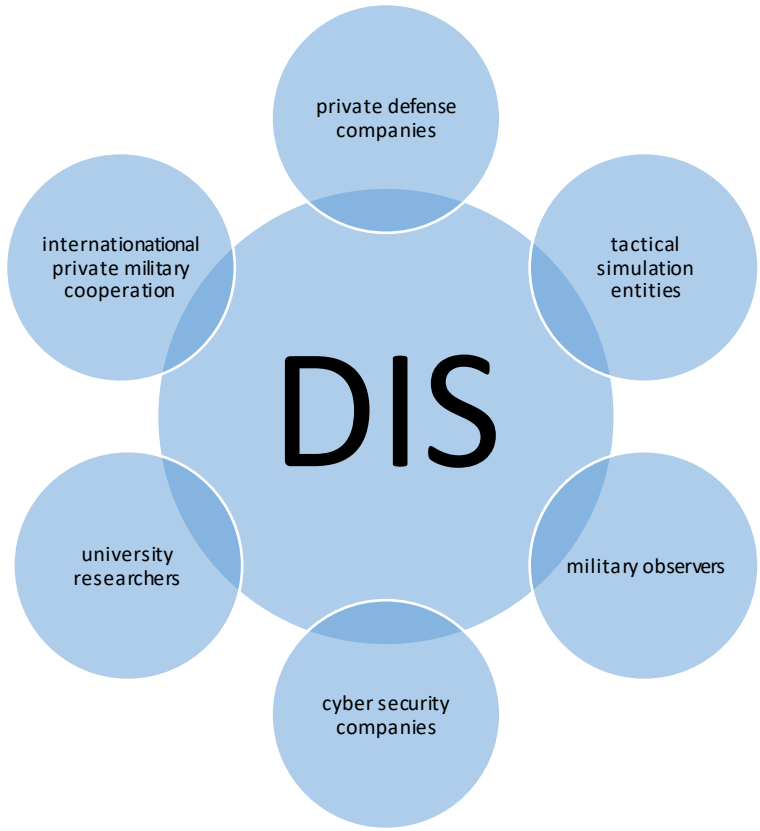


Fig. 1. Aspects of the security sector in which the "SVR"/ DIS can recruit associates

- Representatives of military-applied sports teams, related to the methods in which tactical or field training is conducted. This type of collaborators would also be inclined to enroll in the voluntary reserve, where the benefit of them is that the service is aware of the attitudes and influence of foreign intelligence organizations in circles with a connection to defense (athlete circles, organizers of military simulation events, representatives of defense forums, etc.). Their motivation speaks mostly of belonging and recognition.

- Research units, independent analysts, defense journals, security research teams, and security and defense development and consulting associations. In many countries, "think tanks" guide public opinion and even influence the country's defense policy. Here, there is potential in using such entities for informational or advisory purposes. The content they create and the audience they have accumulated would be in tune with the interests of the service. The motivation here mostly speaks of recognition and finances, and as a level – even through the voluntary reserve, mostly as equal partners in a specific project.

- Shooting clubs and complexes. Here, attention is usually paid to the leadership, but the truth lies in the personnel or staff. Employees of such organizations would be motivated by a need for belonging to an organization and recognition. They turn out to be excited by foreign and domestic politics, defense and international relations. Quite a few of the instructors are also in contact with foreigners, especially representatives of foreign paramilitary units (4).

- External experts, having contact with foreign power seminars, events, courses, etc. These would be people in contact with educational or army organizations on the territory of other countries. Especially valuable as partners, because they would hardly place themselves in dependence on the service, as appointees. Their motivation would be recognition and finances.

- Founders of companies in the field of cybersecurity. Cyber operations would be intriguing for the service, but such people would be motivated mostly by the financial dimensions of cooperation with the SVR, which speaks of a desired independence, in relation to the legal relations with the institution.

The remaining possible collaborators fall into one of the listed categories, so there is no need to additionally list specific activities or external organizations yet.

4. Conclusion

Considering what has been described so far, it is clear that it‘s time to introduce some changes in the methods of hiring personnel, their forms of training and the desired contingent of staff. By using the opportunities offered by the voluntary reserve of the country's armed forces to a fuller extent, the Military Intelligence Service acquires more profound capabilities, and it can also be noted that it contributes to upgrading the general average level of servicemen in the Bulgarian Army.

References

1. Law on the Military Intelligence Service, Republic of Bulgaria
2. Regulations on the Organization and Activities of the Military Intelligence Service, Republic of Bulgaria
3. Stoyanov, S. The Ghostbuster’s art. National Security Magazine, Issue 23, ISSN: 2682-941X & ISSN: 2682-9983, Sofia, 2025
4. Daily protocols of the tactical training complex and instructor’s councils in Sofia region of the Republic of Bulgaria, Sofia, 2025
5. Kaleeva, T. Enhancement of critical infrastructure resilience, Security Science Journal, 6(2):218-232, DOI: 10.37458/ssj.6.2.10, 2025
6. Gechkova, T. The Prioritization of the Measures for the Protection of the National Critical Infrastructure (Strategic Emphases). Economic Alternatives, 31(2).68-76. 2017, Sofia
7. Panevski, V. Critical Infrastructure Resilience Assessment as a Source of Requirements for the Security Systems Development. Proceedings of the Bulgarian Academy of Sciences, 77(12). 1801-1807, 2024, Sofia
8. Law on Defense and the Armed Forces of the Republic of Bulgaria
9. Law on the Civil Servant of the Republic of Bulgaria.

Corruption Risk Assessment in Public Procurement

Stela Stanislavova

University of National and World Economy, Sofia, Bulgaria
sstanislavova12@unwe.bg

Abstract. The paper is focused on studying the corruption risk assessment in public procurements. In order to do so, first of all it aims to elaborate a definition of corruption in public procurement. Second, it outlines some of the key corruption indicators in public procurement. Lastly, but not of importance, the paper aims to evaluate the corruption risk in public procurement in Bulgaria and its implications on the overall corruption index in the country.

Keywords. Corruption, risk assessment, public procurement, red flags, key corruption risk indicators.

1. Introduction

Corruption is a phenomenon well known since Ancient times, however today, due to the state failure, market deregulation, the development of technology and communications, increased competition, globalization, etc., it is presents serious challenge and threat to national security. This negative impact is also highlighted as a major problem in the Updated National Security Strategy of the Republic of Bulgaria, which clearly states that one of the main interests of the state is the prevention and counteraction of corruption and organized crime. Therefore, the relevance of the topic of the article is undeniable, especially taking into consideration the data that Bulgaria is one of the most corrupt countries in the EU, while significant contribution to this is the existing risk of corruption in public procurement.

The main purpose of this paper is to study the corruption risk assessment in public procurement. In order to do so, the paper first of all aims to elaborate a definition of corruption in public procurement. Second, it outlines some of the key corruption indicators in public procurement. Lastly, but not of importance, the paper aims to evaluate the corruption risk in public procurement in Bulgaria and its implications on the overall corruption index in the country.

2. Definition of Corruption in Public Procurement

Corruption is a complex phenomenon with various manifestations. On one hand, it undermines democratic institutions (including electoral process, the rule of law, the stability of government, etc.), and on the other it has a negative impact on economic development. Therefore, it can be analyzed from a legal, social, institutional or economic point of view. First of all, from a legal point of view, corruption is a deviation from legal norms for the sake of personal gain [1]. It is an abuse of powers granted to a certain person by law or contract, the purpose of which is being personal or corporate benefit [2].

From a social point of view, corruption is seen as a betrayal of public interests for personal or group benefits. In other words, corruption is not only a crime, but it is also a negative social phenomenon. It has two dimensions - on the one hand, it is an illegal activity, and on the other, it is an immoral social phenomenon, expressed in a collision between private

and public interest [3]. A similar position on the essence of corruption is supported by Elena Rumyantseva, who in her monograph "Mechanisms for Counteracting Corruption" from 2016 indicates that it is anti-state self-serving behavior, which is characterized by anti-social behavior (abuse of public position, power resources, etc.), in order to pursue a certain self-serving goal (economic benefit, obtaining power advantages, access to resources, information, public procurement, etc.) [4].

From institutional point of view, corruption represents a merger between the economic interests of the authorities and the economic interests of organized crime [5]. Once again this approach discloses that corruption is a means of abusing public power, as a result of which certain personal goals/personal benefit are achieved. By doing so, the corrupt elite places the personal interests above the public ones. Therefore, people who should implement and protect the public interest, for various reasons (mostly financial gain) choose to ignore the general interest at the expense of the private. However, this is a violation of their obligations. In this regard, corruption can have different manifestations. It may take the form of bribery, abuse of power, etc., but the common feature that makes the phenomenon particularly dangerous to society is that under certain circumstances, it is a means of concealing and protecting organized crime, and its organized manifestations are part of organized crime itself.

Last, but not least, corruption can be defined as a threat to economic security, because it is a prerequisite for the formation of inefficiency in the production of public goods by the national economy [6]. It stems from the fact that corruption is possible when public official with discretionary power provide access to or restrict access to certain services, financial resources, etc. to certain companies, individuals, thereby contributing to the uneven distribution of public goods to companies, which can lead to the creation of monopolies, market failure, state failure, etc.

In summary, it can be stated that corruption is a multidimensional phenomenon that encompasses all spheres of life - economy, politics, healthcare, education, etc.; a crime that is expressed in many forms of manifestation (bribery, abuse, trading in influence, etc.); a phenomenon that has a strong negative impact on national security in general, on economic and political stability in particular, and last but not least on society, ethics and morality.

Such definition however is rather obscure when it comes to define corruption in public procurement. Therefore, for the purpose of this article, we define corruption in public procurement as an activity aiming to achieve personal (most often financial) goal, achieved mainly by steering the contract to the favored bidder without detection [7]. Based on this definition, we can derive some important conclusions. First, corruption in public procurement usually includes high-level politicians and business persons, as well as it involves a large amount of public funds [8]. Second, this type of corruption utilizes various ways to achieve its goals, but the two most common ones are using direct contracts awards, thus avoiding competition and benefiting certain companies or proceeding with public procurement, but implementing non-competitive procedures (including sharing inside information with a bidder or tailoring the procurement specifications to fit specific bidder). Considering the risks for the economic and political stability this type of behavior can have, it is rather important to outline the key indicators for corruption risk assessment in public procurement, so that it is not only easier to detect corruption, but to implement preventive measures to prevent or at least minimize the risk of such illegal practices in public procurement.

3. Key Corruption risk indicators in public procurement

Corruption can manifest during the entire procurement process, however sometimes, the process itself can be tainted even before its initial stages due to preceding the procurement legislation aiming to adjust the existing regulation in order to benefit certain parties. Despite

being important, this article aims to study the existing key indicators for corruption during the procurement process, therefore the analysis will focus solely on the corruption schemes applicable on the different stages of the procurement process alone.

During the procurement process, we can identify general and specific red flags regarding the possible existence of corruption schemes. The *general red flag* can be related to:

1) the specifics of the bidding itself – here a possible indicator for corruption can be the single bidding (one offer received despite the market is rather competitive); suspicious bidders (offers received by "shell" companies, in order to create the illusion of competitiveness); suspicious bidding patterns (different companies offers similar bid documents, offers with identical unit prices, a losing bidder becomes a subcontractor, etc.) [9].

2) winning or losing patterns in public procurement – an indicator for a possible corruption can be the repeated win or repeated loss of a company.

3) general red flags for corruption are also the implemented practices to split large procurement into many small contracts, when the procurement process does not develop in a planned or pre-agreed steps or when viable bids are rejected without good reasons [9].

Despite, the so called general red flags can be an important indicator that there is a risk of corruption activities, proving it requires the introduction of a composite indicator that can weight several red flags [10], analysing the probability of corruption practices across the key stages of procurement process.

The first stage of a procurement process is the so called *pre-tender phase*. It involves the assessment of the needs, the drafting of the specifications of the procurement, as well as the selection of the procurement method [11]. Hence, on this stage the main *specific red flags* for corruption practices are as follows:

1) The needs assessment has been manipulated – it can be either artificially induced or inflated, in order to serve the interest of certain bidders [12].

2) Another significant red flag at this stage is when there are indicators that the technical product specifications for a contract have been tailored to either limit the possible bidders or in certain cases are written in a manner that only one company can fulfil them.

3) It is not rare, corruption in public procurement to include using biased eligibility requirements, for example applying restrictive criteria with regards to the technical or financial capacity of the company [8]. Although such criteria a necessary in order to assure that the winning company can fulfil its obligations, it is rather easy sometimes to notice that these requirements are written in a way that undermine the fair competitiveness in order to favour certain company or companies for the given public procurement.

4) Fourth, but not lastly, on this stage the public officials have to select the procurement method. Indicators for corruption at this stage are both splitting the contracts, thus avoiding the legal requirement to apply open tender processes or abusing the so called "urgent" procurement in crisis satiation, leading to increased risk of corruption by avoiding the democratic rule of fair economic competition.

The second phase of public procurement involves the *tender and bidding phase*. On this phase, there are a number of indicators that signal for a possible corruption practices, but for the purpose of this article, they can be systematized in two major groups:

The first group includes indicators for corruption based on certain activities undertaken by the public officials who prepare and conduct the public procurement. These activities mainly aim to limit the access to information about the given procurement to the public, thus allowing the preliminary chosen companies to have advantage in the process. More specifically the specific red flags that often indicate corruption in regards to this group include:

- Shortened or restricted advertising regarding the given public procurement. By limiting the period of the advertisement or restricting its publication to certain location, certain

companies can achieve advantage and thus allows them to be better prepared to participate in the bidding, while preventing the unaffiliated bidders from developing suitable offers in time [13]. In order to mitigate the negative effects of shortened and restricted advertisement, the EU has adopted strict regulations about the transparency of public procurements, explicitly stipulated in Section 2, articled 48-55 of the Directive 2014/24/EU of the European Parliament and of the Council of 26 February 2014 on public procurement and repealing Directive 2004/18/EC [14].

- Despite all efforts for transparency and equal opportunities, public procurement sometimes is related to confidential information leakage. In other words, corruption stems from the possibility a favourable bidder to receive inside information not only about the tender before its publication, but also vital information about the specific specifications, the maximum budget, etc. While, for the above mention red flag there is an effective external control, the control on these informal information flows still remains rather difficult [8].

The second group includes red flags regarding the opportunities for corruption in public procurement that are evaluated based on certain activities undertaken by the bidders itself. There are two major subgroups in this regard, namely:

- In order to manipulate the public opinion and evade one of the general red flags, some companies use the corruption method for dummy bids. In other words, using shell companies or subsidiaries, a favoured bidder submits several bids, thus avoiding the problem of the single bidding.
- Another red flag for corruption in public procurement is the rigging of the bid. The specific indicators that may manifest corruption are coordination between bidders visible in identical bid process, unusual formatting or rotation list of winners and losers in public procurements [15].

A number of red flags can also be discovered during the third phase of the public procurement – *the bid evaluation and ward phase*. Similarly, to the previous phase, there are indicators that can be related predominantly to the bidders or the contracting authorities:

1) Indicators for corruption with respect the activities of the contracting authorities are:

- Rejections of viable bids, without a justifiable reason, usually in order a favoured bidder to win, while a viable lower-cost bidder is rejected on the grounds on minor technicalities.
- Another indicator for corruption can be the implementation of changes in the evaluation criteria, as well as changes in the contract terms and value. Although certain changes in conditions may require minor changes, these should not affect the transparency and the equal opportunities for the bidders. Therefore, any changes, especially ones after the final tender invitations and ones that disadvantage any bidder are strong indicators for a possible corruption practices. Similarly, any changes in contract terms and value, are indicator that may manifest high risk of corruption. These activities quite often may indicate the existence of conflict of interests, where these changes are made, because the decision-maker on behalf of the contracting authorities have undisclosed personal or financial relationships with the winning bidder [13].

2) An indicator for corruption, with respect the activities of the bidder alone, is the use of the so called “bait and switch” tactics. In reality, these tactics when used deliberately, undermine competitiveness, because the bidder wins by offering superior goods, but in reality the consumer receives substitutes of lower quality [16], while the scoring process of the bidding has been undermined.

The *final phase* of public procurement is *contract management*. Several red flags can be listed as potential indicators for corruption. On one hand, indicators for possible corruption are practices such as subcontracting, leading to increased prices of the contract or substandard

performance, due to low quality services or goods. On the other hands, a clear indicator for corruption are also unjustified increase in the total value of the contract. Despite, inflation and rise in cost of certain services and goods can contribute to the need to modify the price of the contract, any other change that is not based on justifiable reasons, is a clear indicator for corruption. Similarly, Lisciandra, Milani and Millemaci in their study conclude that higher expenditures could be attributed to both inefficiencies and corruption [17].

The listed above indicators are not exclusive, however they form a framework that can be useful for the identification and the prevention of corruption schemes in the public procurements. Considering the impact of corruption in Bulgaria, the in-depth knowledge of this framework and its implementation might be used to mitigate the corruption risk in public procurement.

4. Evaluation of corruption risk in public procurement in Bulgaria

Taking into consideration that public procurement is approximately 14% of the GDP in Bulgaria, the evaluation of the corruption risk in this sector is extremely important. This is mainly based on the facts that competitive public procurement practices can boost economic growth by increasing both the efficiency of resource allocations across companies and the productivity growth [18].

Despite this vast importance both for the economic growth and political stability, there is a growing concern that instead of significant decline in corruption practices in public procurement in Bulgaria, there are clear indications for their increase. Undoubtedly, example that confirms these suggestions is the constant growth in single bidding rates for the period 2011-2021 in Bulgaria in a research, prepared as part of the Bulgaria Country Economic Memorandum in 2023.

According to the above mentioned research, since 2011 to 2021, there is a constant growth in the annual average single bidding rates per year in Bulgaria (see Fig. 1). Compared to other EEA countries for the same period, unfortunately our country has the highest single bidding rate in 2021, despite having a relatively lower single bidding rate at 2011 [18]. This is an indication, that despite the implemented stricter rules on public procurement in the EU, since 2017 in Bulgaria faces systematic problems with corruption with regards to public procurement.

In addition, corruption risk in public procurement in Bulgaria is characterized by tailored tender specifications, strong political favouritism, rather visible in large-scale infrastructure projects, and not the least weak oversights over public procurement, leading to increased cases of fraud. Although in recent years the country due to the implementation of the European legislation the lack of transparency has been resolved to some extent, given the strong ties between businesses and political figures, public procurement suffers from this significant exposure to corruption schemes. Thus, businesses with connections to state officials show higher chances of winning contracts, often because of confidential information leakage, single-bid procedures or splitting the large procurement into smaller contracts.

Considering the existing level of corruption in public procurement in Bulgaria, there a number of obligatory steps to be undertaken for mitigating the effects of corruption. The easier detection of red flags across the country require digitalization and increased transparency along with improved mechanisms for reporting corruption and unfair practices and more efficient and effective monitoring and accountability – especially targeting the high-levels of corruption.

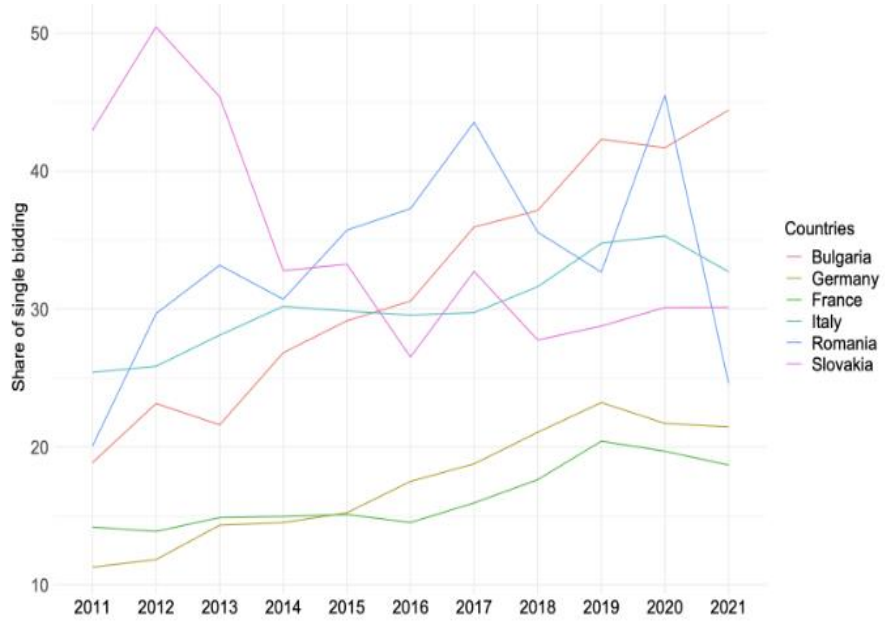


Fig. 1. Annual average single bidding rates per year, selected EEA countries, 2011-2021 [Source: Fazekas, Mihaly, Poltoratskaia, Viktoriia, Toth, Bence. Corruption risks and state capture in Bulgarian Public Procurement. Washington: World Bank Group, 2023 (https://openknowledge.worldbank.org/server/api/core/bitstreams/b512907f-4b38-441c-8d77-4c2ed97c684c/content)]

5. Conclusion

Public procurement is one of the spheres of public governing that is usually characterized by higher risk of corruption, although its levels may vary significantly in the different countries. The analysis in this paper outlined some of the most frequent red flags that can indicate corruption practices, thus showing the many forms it can take. Based on these indicators, corruption risk can be easily assessed and thereafter the conclusions can be used to address the most problematic areas.

However, given the case in Bulgaria, despite the persistent demand from the civil society for the implementation of anti-corruption reforms, most of the efforts to mitigate the corruption practices in public procurement remain ineffective. The main cause for this is the existing institutional weakness and political unwillingness to act, stemming from the state capture and political favourism.

References

1. Penev, Yordan. Corruption – present, past, prevention. Bulletin of the Faculty of Police, AMVR, issue 29, 2012 [Penev, Ĭordan. Koruptsiyata – nastoyashte, minalo, preventsiya. Byuletin na fakultet „Politsiya“ AMVR, br. 29, 2012]
2. Stankov, Boyan. Criminology: Types of Crime. Varna: VFU "Chernorizets Hrabar", 2003 [Stankov, Boyan. Kriminologiya: vidove prestŭpnost. Varna: VSU „Chernorizets Hhrabŭr“, 2003].
3. Shopova, Petya. Corruption and anti-corruption in Bulgaria. Counteraction to corruption in Bulgaria. Ruse: Academic publishing house - University of Ruse, 2019 [Shopova, Petya. Koruptsiya i

antikoruptsiya v Bulgaria. Protivodeĭstvie na koruptsiyata v Bulgaria. Ruse: Akademichno izdatelstvo - Rusenski universitet“, 2019].
4. Rumyantseva, Elena. Mechanisms of countering corruption. Moscow: Direct-Media, 2016.
5. Velkova, Lydia. Corruption as a threat to economic security. Sofia: DioMira, 2015 [Velkova, Lidiya. Koruptsiyata zaplakha za ikonomicheskata sigurnost. Sofia: DioMira, 2015].
6. Deliverski, Jordan. Corruption and Prevention of Corruption. Sofia: About Letters – about writing, 2015 [Deliverski, Jordan. Koruptsiya i preventsiya na koruptsiyata. Sofia: Za bukvite – o pismenekh, 2015].
7. World Bank: Fraud and corruption awareness handbook: how it works and what to look for. A handbook for staff. Washington, D.C.: World Bank Group, 2010 (https://documents1.worldbank.org/curated/en/100851468321288111/pdf/575040WP0Box351Corruption1Awareness.pdf)
8. Kofrán, Daniel, Czibik, Agnes, Fazekas, Mihaly. Corruption Risk Indicators in Public Procurement: An extended opentender.eu framework. Risk assessment methodology for the iMonitor project, 2024 (https://imonitor.govtransparency.eu/wp-content/uploads/2024/03/D2.2-Risk-assessment-methodology.pdf)
9. World Bank: Warning signs of fraud and corruption in procurement. Washington D.C.: World Bank Group, 2020 (https://documents1.worldbank.org/curated/en/223241573576857116/pdf/Warning-Signs-of-Fraud-and-Corruption-in-Procurement.pdf)
10. Fazekas, Mihály, Kocsis, Gábor. Uncovering High-Level Corruption: Cross-National Objective Corruption Risk Indicators Using Public Procurement Data. British Journal of Political Science, vol. 50 (1), 2020, pp. 155–164.
11. Kohler, Julian, Dimacesco, Deirdre. The risk of corruption in public pharmaceutical procurement: how anti-corruption, transparency and accountability measure may reduce this risk. Global Health Action. Vol. 13 (1), 2020 (https://pmc.ncbi.nlm.nih.gov/articles/PMC7170361/)
12. Morgner, Matthias, Chene, Marie. Corruption risks in public procurement. Transparency International, 2015 (https://knowledgehub.transparency.org/guide/topic-guide-on-public-procurement/4899)
13. Gnaldi, Michela, Del Sarto, Simone. Validating Corruption Risk Measures: A Key Step to Monitoring SDG Progress. Social Indicators Research, vol. 175, pp. 1045–1071.
14. Directive 2014/24/EU of the European Parliament and of the Council of 26 February 2014 on public procurement and repealing Directive 2004/18/EC (https://eur-lex.europa.eu/eli/dir/2014/24/oj/eng)
15. International Anti-Corruption Resource Center: Red flags : Unusual bidding patterns, 2026 (https://guide.iacrc.org/detection/red-flags-of-corruption-bid-rigging-and-other-schemes/red-flag-unusual-bidding-patterns/)
16. Office of Procurement Ombud: “Bait and switch”: Procurement practice review of replacement resources in professional services contracts. 2025 (https://opo-boa.gc.ca/praapp-prorev/2025/ep-rc-10-2025-eng.html)
17. Maurizio Lisciandra, Riccardo Milani, Emanuele Millemaci. A corruption risk indicator for public procurement. European Journal of Political Economy, vol. 73, 2022,102141, (https://www.sciencedirect.com/science/article/abs/pii/S0176268021001166)
18. Fazekas, Mihaly, Poltoratskaia, Viktoriia, Toth, Bence. Corruption risks and state capture in Bulgarian Public Procurement. Washington: World Bank Group, 2023 (https://openknowledge.worldbank.org/server/api/core/bitstreams/b512907f-4b38-441c-8d77-4c2ed97c684c/content)

An Analytical Model for Assessing Corruption Risk in Public Procurement: Evidence from Sensitive Sectors

Stela Stanislavova

University of National and World Economy, Sofia, Bulgaria
sstanislavova12@unwe.bg

Abstract. Public procurement is persistently exposed to elevated corruption-related risks, particularly in sectors characterised by strategic importance, limited transparency, and constrained competition. Conventional governance approaches rely predominantly on ex post controls, which often fail to capture structural and procedural vulnerabilities embedded within procurement systems. This article adopts a risk-oriented perspective, conceptualising corruption not as a series of isolated violations but as an outcome of institutional design, procedural discretion, and governance arrangements. An analytical, indicator-based model for assessing corruption risk in public procurement is proposed, with a specific focus on sensitive sectors such as defence and security. The model estimates relative risk exposure across multiple governance domains, supporting ex ante assessment and preventive decision-making rather than enforcement or attribution of wrongdoing. Illustrative analytical scenarios demonstrate how procedural configurations, market conditions, and oversight mechanisms interact to shape differentiated risk profiles. The findings suggest that structured analytical models can enhance integrity and accountability in public procurement when applied as decision-support tools that complement existing regulatory, auditing, and investigative frameworks, particularly in environments where transparency is structurally constrained.

Keywords. Public procurement; corruption risk; risk assessment model; sensitive sectors; governance and control; ex ante analysis.

1. Introduction

The importance of modern systems of public governance stems from their complexity and the significant amounts of public money involved. Public governance systems around the world have been shown to vary in the levels of control and accountability they provide, as well as in their ability to procure goods and services quickly and efficiently. Despite the detailed regulations in place to guide public governance systems, public governance is still viewed as vulnerable, partly because formal compliance does not always translate into robust integrity outcomes [1]. It is apparent that the complexity of these governance systems has few, if any, instances of unlawful acts, yet remains vulnerable to a lack of accountability, unexercised control, and gaps across all parts of the system, collectively termed ‘the system’.

The lack of clearly defined control mechanisms across the numerous parts and stages of the winding and complex systems of public governance defines the system as a network of potential vulnerabilities. The numerous roles, layers, and mechanisms through which control is exercised may coexist without sufficient coordination, reducing their effectiveness. Overall, public governance systems create conditions in which discretion may exceed oversight capacity, thereby increasing opportunities for unaccounted decisions that undermine integrity and efficiency.

Among the most prominent issues are the pronounced obstacles to transparency and the strategic importance of the defence and security sector. This procurement domain involves the classification of documents, a limited number of supplier competitors, and unique, highly nationalistic legislation justified on the grounds of national security [5]. While these factors do not necessarily indicate a higher prevalence of corruption, they generate a greater number of risk indicators that complicate the application of transparency and accountability measures. Regulatory instruments, which may be overly broad, remain ineffective, particularly when control and oversight are constrained by inadequate information or a deficient institutional framework.

The gap between formal regulations and political practice underscores the need to consider corruption risk in a more systematic and proactive manner. Most control efforts have focused on post hoc analysis of irregularities through audits, investigations, or court actions. These measures are fundamental but reactive, tending to address only a small part of the underlying deficiencies. Increasingly, the focus is shifting to risk analysis and other ex ante tools that facilitate the identification of a risk environment before an imbalance occurs, emphasising prevention rather than punishment [7].

Given this context, this article outlines an analytical model of corruption risk assessment in public procurement, focusing on sensitive sectors. The model is based on standardised corruption risk assessment indicators informed by the state of institutional arrangements, the design of the procedures, situational factors, and the absence of corruption risk at the individual level. It does not aim to establish the presence or absence of corruption, but to assist in relational risk assessment and risk profile decision-making. This article examines high-sensitivity procurement as a way of contributing to the analytical and evidence-based discourse on risk management in governance, transparency, and accountability systems in public procurement.

2. Public Procurement and Corruption Risk

In the academic literature, the risk of corruption in public procurement is not treated as a uniform or singular issue but is recognised as a multi faceted phenomenon reflecting various governance vulnerabilities in procurement systems [1]. Public procurement is intrinsically vulnerable to integrity risks due to the high volume of transactions, the large financial stakes involved, and multiple interaction points between public officials and private suppliers, which collectively create opportunities for unethical practices such as bid rigging, bribery, and conflict of interest. Indicators widely used to detect corruption risk include red flags such as lack of transparency, limited competition, irregular bidding patterns, and excessive use of discretionary contract awards [5].

Legally deficient procurement structures refer to deviations from established rules and procedures that are empirically verifiable through audits or judicial review. Yet, even where formal compliance appears to be present, ethical ambiguities can persist — for example, when procurement specifications are tailored to favour certain suppliers, reducing fairness and neutrality in the process. Such practices may not violate legal requirements but nonetheless reflect a poorly designed system that allows discretion without adequate constraints, thus increasing corruption risk.

Systemic risk exposure denotes a situation in which the very design of procurement processes, organisational arrangements, and oversight mechanisms makes a system susceptible to corruption, even if no explicit violation is evident [7]. Public procurement processes comprise multiple interconnected stages, starting with needs assessment, proceeding to specification drafting, competitive tendering, contract execution, and ending with post award monitoring. Each stage can generate distinct risk profiles. For example, the assessment of

needs may be influenced by undue external pressures, and drafting specifications may incorporate unnecessary technical detail to favour certain market actors.

Discretionary power tends to be most acute when verification capacity is low. Rules designed to ensure uniformity and consistency may not be evenly or consistently applied, especially where complex technical requirements or expedited timelines prevail. This can create gaps in oversight and make post hoc explanations for decisions difficult for external auditors or oversight bodies. Consequently, focusing exclusively on formal compliance overlooks informal, relational, and procedural corruption, which often goes undetected by compliance based models.

Improving risk assessment therefore requires moving beyond binary legal compliance and towards frameworks that identify and measure patterns and signals that indicate elevated risk exposure. Data driven and red flag indicator approaches aim to capture patterns such as collusion, repeated contract awards to the same suppliers, unusual bid price distributions, or opaque supplier relationships, offering more nuanced insights into corruption risks in public procurement [8]. This shift towards integrated, multi dimensional risk frameworks supports both preventive governance and better allocation of oversight resources.

3. Sensitive Sectors as High-Risk Environments

Various forms of public procurement have been identified in the literature as carrying the highest risk of corruption, and defence and security procurement continues to rank among these forms due to its intrinsic characteristics such as confidentiality requirements, market concentration, and legislative exemptions. Analyses by Transparency International and other governance bodies note that defence procurement often involves classified information and limited visibility, which can impede transparency and oversight mechanisms designed to safeguard integrity [4].

Strategic confidentiality is a primary factor. In defence and security procurement, contracts frequently involve classified or sensitive details related to military capabilities, operational readiness, or advanced technology. Although such confidentiality may be justified on security grounds, it structurally limits public access to procurement processes and reduces opportunities for external scrutiny. The absence of transparency in these contexts is associated with heightened corruption risk because it removes or weakens key accountability mechanisms.

Another risk arises from market concentration and high entry barriers. Defense and security markets are generally characterised by a small number of specialised firms with long term relationships with state actors. High sunk costs, technical specificity, and limited substitute suppliers can reduce competition and strengthen positional advantages for incumbent firms. Competitive dynamics that favour single bidder procedures or repeated awards to the same suppliers can weaken safeguards that would otherwise mitigate the risk of collusion or rent seeking behaviour.

Procurement frameworks and legislative derogations further complicate the governance environment. Defence procurement often operates under special legal regimes that allow deviations from standard public procurement rules, such as flexibility in advertising requirements, timing, or competitive procedures. Although these provisions are often justified to address national security needs, the broader discretion they afford increases the potential for unmonitored decision making and weakens consistency in application.

The risk profile in defence and security procurement should not be interpreted as evidence of inherently higher corruption prevalence; rather, it reflects structural constraints that challenge traditional oversight mechanisms [7]. The literature indicates that strategic

confidentiality combined with limited competition and procedural flexibility can create governance environments where standard transparency based controls are less effective.

Recent institutional responses emphasise integrity building and risk oriented governance frameworks that do not rely solely on access to information but aim to strengthen internal safeguards, promote accountability, and build organisational capacity to handle corruption exposures in sensitive contexts [9]. Among these, NATO's Building Integrity Programme provides a structured framework for strengthening transparency and accountability in defence institutions, offering tailored tools such as self-assessment and peer review processes to mitigate corruption risks even where confidentiality constraints limit traditional oversight [12].

In this context, defence and security procurement provides a rich setting for evaluating analytical risk frameworks. The structural limitations inherent in these sectors highlight the challenges of applying compliance centric models and emphasise the value of structured, indicator driven approaches that can systematically assess risk even when transparency is limited.

4. Methodological Approach and Data Logic

The analytical framework presented in this paper conceptualises the assessment of corruption risk in public procurement as a comparative and risk oriented exercise rather than a predictive or purely diagnostic system [2]. Corruption cannot be reliably measured using solely quantitative techniques, nor can it be determined from one or many disparate measures. Instead, the framework seeks to understand relative risk exposure by evaluating the interplay of institutional arrangements, procedural discretion, and context-specific constraints that influence the likelihood of unaccounted-for or other-regarding preferences.

Accordingly, the framework does not attempt to identify individual acts of corruption or perpetrators. Its objective is more modest yet practical: to provide ex ante evaluation of procurement processes or environments with elevated corruption risks, serving as a decision-support tool that complements formal legislative, auditing, and investigative systems [5]. By focusing on relative risk rather than absolute status, the model allows for comparative analysis across procurement contexts while recognising the inherent limits of reputation-based or formal assessments.

From a data perspective, the model employs a mixed-methods approach, integrating qualitative and quantitative components. Qualitative input is derived from institutional reviews, governance frameworks, and expert assessments that describe risk patterns in public procurement. These inform the construction of indicators designed to encapsulate governance phenomena rather than solely legal compliance. Quantitative operationalisation enables these indicators to be measured, counted, and compared across different procurement practices, providing structured guidance for risk scoring and prioritisation.

In sensitive procurement domains, data are often incomplete, restricted, or fragmented. To address this, the model utilises proxy indicators, recognising their limitations but leveraging them to reflect structural and systemic risks. Proxy indicators should be interpreted within the collective context of all indicators: the more interdependent and mutually supportive the indicators, the higher the inferred risk exposure.

The framework also emphasises methodological transparency. Users are guided on how to derive segmented risk profiles from prior assumptions and how to operationalise these in terms of indicators [11]. The design highlights the methodology, intended outcomes, and means to achieve them, demonstrating applicability across institutions with varying governance frameworks. Importantly, the model is not intended as a tool of positive science,

producing universal truths; rather, it is a pragmatic instrument to address specific governance challenges, acknowledging inherent ambiguities and limitations.

5. Development of the Analytical Corruption Risk Model

The analytical model developed in this article is conceived as a multi-dimensional framework that captures the complexity of corruption risk in public procurement without reducing it to a single explanatory factor [2]. The central premise is that corruption risk arises from the interaction of multiple governance dimensions, each of which may appear unproblematic in isolation yet, together, create conditions of heightened vulnerability. For this reason, the model is structured around a set of interrelated risk domains rather than a linear causal chain.

The first domain concerns the institutional context in which procurement occurs. This domain reflects the broader governance environment, including the stability of institutional arrangements, the clarity of role allocation, and the capacity of oversight bodies. Institutional fragmentation, overlapping competences, or weak coordination mechanisms may increase ambiguity in decision-making. Such ambiguity does not automatically lead to malpractice, but it can dilute accountability and make it more difficult to attribute responsibility ex post. Indicators in this domain are therefore selected to approximate institutional coherence and control capacity, rather than legal compliance alone [6].

The second domain focuses on procedural design, capturing how procurement rules are operationalised in practice. This includes the choice of procurement procedure, the level of discretion afforded to contracting authorities, and the degree of standardisation across comparable contracts. Procedures that allow extensive negotiation, accelerated timelines, or deviations from open competition may be justified by operational constraints, particularly in sensitive sectors. However, these features tend to reduce verifiability and increase reliance on expert judgement, which, when combined with limited oversight, may elevate risk [8]. Indicators in this domain aim to reflect procedural flexibility and transparency, recognising that flexibility itself is not inherently problematic.

The third domain concerns market conditions, particularly the structure and competitiveness of supplier markets. Market concentration, barriers to entry, and patterns of repeated contract awards are relevant here. In procurement environments where only a small number of suppliers are technically capable of delivering specific goods or services, competition may be structurally constrained. While such constraints are often unavoidable, they may weaken one of the principal safeguards embedded in procurement systems. Indicators in this domain aim to capture the extent to which competitive pressure can realistically operate, rather than assuming an idealised market structure [4].

The fourth domain concerns control and oversight mechanisms, encompassing both internal and external scrutiny. This includes ex ante controls, such as approval hierarchies and documentation requirements, as well as ex post mechanisms, such as audits and performance reviews. The effectiveness of these mechanisms depends not only on their formal existence but also on their timing, independence, and access to information. Where controls are largely procedural or delayed, their capacity to mitigate emerging risks may be limited. Indicators in this domain therefore focus on control intensity and coverage across the procurement lifecycle.

Indicator selection across these domains follows a two-step process. First, indicators are derived from recurring risk factors identified in the literature and in governance practice. Second, they are operationalised to allow comparison across cases, even when data availability varies. The model deliberately avoids highly granular indicators that would require sensitive or inaccessible information. Instead, it prioritises indicators that can be plausibly observed or approximated in constrained transparency environments.

Weighting and aggregation are treated as analytical choices rather than objective truths. Indicators within each domain are combined to generate domain-specific risk scores, which are then aggregated into an overall risk profile. Weighting schemes may vary according to policy priorities or sectoral characteristics, and the model is designed to accommodate this variation. Importantly, aggregation does not produce a definitive "high" or "low" corruption risk classification. Rather, it yields a relative positioning that highlights patterns of vulnerability.

To illustrate, consider a procurement procedure conducted under accelerated timelines, with a restricted number of bidders and limited public disclosure due to security considerations. Individually, none of these features necessarily indicates impropriety. However, when they coincide—particularly in the absence of strong compensatory controls—the combined risk profile may warrant closer examination. The model captures this interaction by allowing multiple indicators across procedural and market domains to reinforce one another.

Throughout its design, the model is explicitly framed as a decision-support tool. It does not replace legal assessment, investigative judgement, or contextual expertise. Instead, it provides a structured way to synthesise diverse signals into an interpretable risk profile. By making assumptions transparent and avoiding binary conclusions, the model aims to support preventive governance while respecting the complexity and sensitivity of public procurement, particularly in high-stakes sectors [2].

6. Application of the Model to Sensitive Procurement Context

As with the previous section, the objective is not to develop empirical generalisations or evaluate particular instances in the real world. Rather, it is to demonstrate the model's value in providing more structured analytical clarity in situations characterised by low transparency and high institutional risk. To achieve this objective, the analysis is framed within hypothetical situations. The case studies present analytical situations within contexts that require a necessary level of secrecy or are otherwise inaccessible due to their classified nature [3].

An illustrative example involves the procurement of specialised defence equipment via closed bidding, on the basis of national security grounds for no-bid procurement. From a compliance perspective, such a procedure may be fully within the bounds of applicable legislation and the relevant sectoral exemption. However, the model seeks to assess risk across multiple domains. In a stable institutional framework, with clear roles and no overstepping of institutional hierarchies, oversight bodies are likely to mitigate such risks. Thus, in this case, the procedural design domain may be characterised by low risk. However, the closed bidding itself may pose high risk due to a lack of competition, reinforcing the negotiated nature of the procurement process. The operational risk is reasonably apportioned within the parameters of such processes, as there are no indications of illegality.

The second hypothetical example concerns procurement undertaken on shorter timeframes due to operational necessity. Time constraints are commonplace in defence and security and may be justified and unavoidable. In the model, time-pressed, streamlined procedures affect the process design domain by eliminating or reducing the levels of place for checking and external supervision. If such processes are coupled with a supplier's market in the benefits and a lack of post-award monitoring, the risk profile may be greater than in instances where the urgency is coupled with compensatory safeguards. On the other hand, the absence of time may be compensated for by stronger internal controls or an independent technical review. In the model, there is also the possibility of examining compensatory governance. Where market concentration is an unavoidable fact of life, the presence of the control and oversight layer may, to some extent, balance the heightened risk indicators of the

market conditions domain. Combining pre-award processes such as documenting decision rationales, retaining a layered approach to approvals, and overseeing post-award performance can improve the risk profile. The model aims not to treat structural constraints as deficiencies, focusing on how institutions utilise such constraints [8].

The model's logic becomes clear through these examples. Risk profiles are assessed relative to one another rather than against an absolute standard. This approach is particularly suitable in sensitive areas where applying uniform transparency thresholds is not feasible. The model supports prioritisation by identifying procurement activities that deviate from established institutional risk patterns, thereby reallocating focus and resources to areas where additional scrutiny is warranted [9].

Model outputs, particularly in the sensitive sectors case, should be interpreted with caution. Higher risk exposure should not be taken as prima facie evidence of wrongdoing, nor should lower risk profiles be taken as assurances of no wrongdoing. The model offers a way to identify governance gaps. In procurement within the defence and security sectors, where conventional transparency tools are limited, such a governance gap analysis can support positive governance in line with the institutional sensitivities of public sector functions [10].

7. Discussion: Strengths, Limitations, and Policy Implications

The analytical model presented in this article offers several contributions to assessing corruption risk in public procurement, particularly in contexts where conventional transparency mechanisms face structural constraints. One of its primary strengths is the transparency of its underlying logic. By explicitly articulating the domains, indicators, and aggregation principles used to construct risk profiles, the model enables users to understand how assessments are formed and to interrogate the assumptions embedded within them. This openness distinguishes the approach from black-box assessments and supports informed interpretation rather than mechanical application [1].

A further strength is the model's adaptability. Although developed with sensitive sectors in mind, the framework is not inherently sector-specific. Its domains—covering institutional context, procedural design, market conditions, and control mechanisms—are sufficiently general to be applied across a wide range of procurement environments. At the same time, the model allows contextual adjustment through indicator selection and weighting, making it responsive to differing legal regimes, market structures, and policy priorities. This flexibility enhances its compatibility with existing control systems, including audit frameworks and internal risk management processes.

At the same time, the model's boundaries warrant careful consideration. Assessing corruption risk remains inherently dependent on data quality and availability. In sensitive procurement contexts, data are often incomplete or uneven, requiring reliance on proxy indicators [6]. While proxies can approximate underlying governance conditions, they may also reflect institutional capacity constraints or policy trade-offs rather than corruption-related vulnerability in a narrow sense. This introduces uncertainty that cannot be fully resolved through analytical refinement alone.

Indicator weighting is another source of subjectivity. Although weighting schemes are made explicit and can be adjusted, they inevitably reflect normative judgements about the relative importance of different risk factors. Different institutional actors may reasonably assign different weights based on their mandates or risk perceptions. The model accommodates this variation, but it cannot eliminate the interpretive dimension of risk assessment.

Perhaps the most significant limitation is the difficulty of capturing informal influence networks and relational dynamics. Many forms of undue influence operate through social ties, reputational dependencies, or long-term relationships that are not readily observable through

formal indicators. Analytical models, by their nature, tend to privilege what can be measured or approximated. There is therefore a risk that overreliance on structured assessments may create a false sense of control, particularly when qualitative judgement and contextual knowledge are marginalised. From this alternative perspective, analytical tools should be viewed as aids to governance, not substitutes for institutional vigilance [11].

The policy implications of the model are therefore framed modestly. Rather than serving as a basis for enforcement or sanctions, the model is best suited to supporting preventive functions. It can inform audit planning by highlighting procurement processes with clustered risk indicators, helping oversight bodies prioritise limited resources. Within contracting authorities, the model may support internal control by drawing attention to procedural configurations that warrant additional safeguards. At an institutional level, it can facilitate self-assessment and comparative learning, allowing organisations to reflect on how structural choices affect risk exposure over time.

In sum, the analytical model's value lies not in its ability to definitively resolve corruption risk, but in its capacity to structure attention and dialogue around governance vulnerabilities. Used critically and in conjunction with existing control mechanisms, it offers a pragmatic contribution to strengthening integrity in public procurement.

8. Conclusion

Public procurement remains one of the most complex and vulnerable domains of public governance, not because corruption is inevitable, but because procurement systems operate at the intersection of discretion, technical complexity, and institutional constraints. As the preceding analysis has suggested, corruption risk in this context is better understood as an emergent property of governance arrangements rather than a series of isolated violations. This perspective helps explain why formal compliance with regulatory frameworks does not always translate into effective risk mitigation, particularly in sectors where transparency is structurally limited.

Against this background, analytical and ex ante approaches to corruption risk assessment are increasingly relevant. Reactive mechanisms, such as audits and investigations, remain indispensable, yet they often address symptoms rather than underlying conditions. The model proposed in this article addresses this gap by offering a structured, indicator-based framework for assessing relative risk exposure across procurement processes. Its purpose is not to identify misconduct or predict outcomes, but to support informed judgement by making governance vulnerabilities more visible and comparable [6].

The model's contribution lies primarily in its conceptual clarity and practical orientation. By organising risk factors into distinct yet interrelated domains, it offers a coherent lens for examining procurement environments. The emphasis on transparency of assumptions, adaptability to context, and compatibility with existing control systems reflects a deliberate effort to balance analytical structure with epistemic caution. Rather than producing definitive classifications, the model encourages reflective interpretation and dialogue among institutional actors.

At the same time, the model's limitations underscore the need for ongoing refinement and empirical testing. Future research may explore more granular indicators, assess the model's performance across different jurisdictions, or examine how weighting schemes shape risk profiles. There is also scope to integrate ICT-based monitoring tools, such as automated data extraction or pattern analysis, to improve the timeliness and consistency of risk assessment [11]. Such developments, however, should be approached carefully to ensure that technological capacity does not eclipse contextual understanding.

Ultimately, strengthening integrity in public procurement is not a matter of implementing a single analytical tool or regulatory fix. It is an ongoing process of institutional learning and adjustment, shaped by evolving risks, policy priorities, and governance capacities. Analytical models, such as the one presented here, can contribute to this process by focusing attention and supporting preventive action, but they remain part of a broader ecosystemof accountability rather than a standalone solution.

References

1. OECD, Integrity in Public Procurement: Good Practice, Paris, France: OECD Publishing, 2007. https://www.oecd.org/en/publications/integrity-in-public-procurement_9789264027510-en.html;
2. OECD, Managing Public Procurement Risks in Romania, Paris, France: OECD Publishing, 2023. https://www.oecd.org/en/publications/managing-public-procurement-risks-in-romania_5dfda60e-en.html;
3. OECD, Fighting Corruption and Promoting Integrity in Public Procurement, Paris, France: OECD Publishing, 2005. https://www.oecd.org/en/publications/fighting-corruption-and-promoting-integrity-in-public-procurement_9789264014008-en.html;
4. Basel Institute on Governance, Preventing Corruption in Public Procurement, Basel, Switzerland: Basel Institute on Governance, 2016;
5. Transparency International, Preventing Corruption in Public Procurement, Berlin, Germany: Transparency International, 2019;
6. Transparency International, Transparency, Accountability and Integrity in Public Procurement Systems, Anti-Corruption Helpdesk Brief, Berlin, Germany: Transparency International, 2024.
7. D. Dávid-Barrett, D. Holmes, and R. Baturo, "Corruption in public procurement: Finding the right indicators," European Journal on Criminal Policy and Research, vol. 23, no. 3, pp. 245–267, 2017. doi: 10.1007/s10610-017-9349-3
8. A. T. Bastos, R. P. Dantas, and L. M. Silva, "Fraud, corruption, and collusion in public procurement activities: A systematic literature review," Applied Network Science, vol. 7, art. no. 83, 2022. doi: 10.1007/s41109-022-00521-5
9. World Bank, Enhancing Government Effectiveness and Transparency: The Fight Against Corruption, Washington, DC, USA: World Bank Group, 2020.
10. World Bank, "Early detection of fraud and corruption in public procurement through technology," Washington, DC, USA: World Bank Group, 2020.
11. G. Bouckaert and J. Halligan, Managing Performance: International Comparisons, London, UK: Routledge, 2019, pp. 112–125.
12. NATO, Building Integrity Policy and Programme, Brussels, Belgium: NATO Building Integrity Programme, 2016–2025.

Key Corruption Risks in Bulgarian Public Procurement

Stela Stanislavova

University of National and World Economy, Sofia, Bulgaria
sstanislavova12@unwe.bg

Abstract. The paper is focused on studying the key corruption risks in Bulgarian public procurement. The first part examines the correlation between corruption and public procurement. The second is focused on the indicator single-bidding in public procurement in Bulgaria, as well as outlining some of the reasons for its high percentage. The third part outlines some of the characteristics of the political favouritism in public procurement and the red flags associated with it.

Keywords. Corruption risk, public procurement, single-bidding, political connections and favoritism, irregularities in procedures

1. Introduction

Public procurement is highly prone to corruption. Therefore, given the fact that each year the purchase of goods, services, and construction services with public funds is estimated on EU average about 16% of the GDP, on one hand this allocation of funds is rather beneficial to the economy, but on the other by creating unfair access to government procurement contracts it not only hinders competitiveness, but also can increase the cost of public services. In any case, corruption in public procurement damages the public interest. By favouring the private (corporate) interest, while most of the times benefiting certain public officials as well.

Unfortunately, despite the number of legislative changes in correspondence to the EU legislation, Bulgarian public procurement system is still considered to be of high risk of corruption. Therefore, it is vital to be able to understand and recognise the key corruption risks in the Bulgarian public procurement, especially taking into consideration the fact that the corruption can have many faces, including by favouring a certain bidder, excluding qualified bidders, modifying terms at the contract implementation stage, etc.

Hence, the main purpose of this article is to examine and outline the key corruption risk in public procurement in Bulgaria, but analyzing the relation public procurement-corruption and examining two of the main corruption indicators – the high number of single-bids and the existing political favouritism.

2. Public procurement – a high corruption risk in Bulgaria

Taking into consideration that the share of public procurement in GDP in Bulgaria varies between 10-14%, it is considered a significant drive for economic growth through innovation, encouraging sustainable practices, boosting local businesses and creating jobs. Although this share of GDP lacks behind the average in EU countries, there is a gradual increase over the years, clearly noticeable in the Table below.

This increase in public investment is also visible in the number of public procurements, that almost doubles from 11 894 in 2014 to 22 736 in 2024 [2]. However, this alone comes with significant risks, mainly due to the fact that public procurement is not only highly susceptible to disruption by corruption [3], but because the Bulgarian procurement system

itself still lacks competition and remains to some extent rigged with political favouritism and limited oversight.

Table 1. General government expenditure on procurement in Bulgaria [1]

Year	Million euro	Percentage of GDP
2020	6 572	10.9
2021	7 377	10.2
2022	8 672	10.1
2023	10 668	11.2
2024	21 576	14.0

Since 2016, the Bulgarian legislation on public procurement and its regulatory regime has been significantly shaped by the EU legislative framework. By implementing the core principles of public procurement procedures, elaborated in Directive 2014/24/EU of the European Parliament and of the Council of 26 February 2014 on public procurement and repealing Directive 2004/18/EC, Directive 2014/23/EU of the European Parliament and of the Council of 26 February 2014 on the award of concession contracts, Directive 2014/25/EU of the European Parliament and of the Council of 26 February 2014 on procurement by entities operating in the water, energy, transport and postal services sectors and repealing Directive 2004/17/EC and Directive 92/13/EEC of 25 February 1992 coordinating the laws, regulations and administrative provisions relating to the application of Community rules on the procurement procedures of entities operating in the water, energy, transport and telecommunications sectors, the Bulgarian regulatory framework made a significant progress. However, the frequent amendments of the Bulgarian Public Procurement Act are indicative of constant struggle to reform the public procurement in a way to reduce the risk of corruption.

This risk is considerable higher to certain sectors and areas of the economy, some due to restricted competition, others because of tailored tender requirements or merely the procurement volumes. Based on this, we can conclude that in Bulgaria the sectors that are more prone to corruption are:

- Construction and infrastructure – a sector that is frequently involved in cases of large-scale corruption, mainly because of over-specified, tailored tender requirements that limit the participation to favoured firms.
- Energy sector – despite its importance, because of the nature of this sector, is carries increased risk of corruption. For example, it can be exemplified by cases like the Lukoil refinery, where non-competitive practices were shielded [4].
- Healthcare sector – in the past years several journalists' investigations have discovered major corruption schemes in this sector. To great extend the increased risk of corruption in this sector is due to the massive procurement volumes for pharmaceuticals and hospital equipment.
- Significant corruption risk is also associated with the EU-funded projects, which are rigged with irregularities, lack of transparency, and administrative errors.

Despite the implemented European regulations on the public procurement regime in Bulgaria, the corruption risk remains relatively high. Based on the specifics of the Bulgarian market and the allocation of public funds to services, construction and deliveries, as well as the wide spread practices, we can conclude that the key corruption risks consists of single-bidding, customized tender specifications, political favouritism and non the least the high share

of negotiated procedures without prior publication, as well as there is also a significant share of procurement done by direct negotiations.

This practices therefore deter business from participating in public procurement, citing biased conditions and evaluations [5], thus in addition deepening the problem with the limited competition. By addressing these weaknesses, according to the World Bank, Bulgaria could potentially save about 5.3% of the total value of the contracts, or approximately BGN 1.3 billion [6].

3. Single bidding as a corruption risk in Bulgarian public procurement market

Fair competition is vital to mitigate the corruption risk in public procurement. However, it is rather difficult to implement, when a large percentage of the procurement contracts are awarded without competition, partly due to "single-bidding", partly because of direct negotiations and negotiated procedures without prior publication. It is worth mentioning that, among them the use of single-bidding is rather problematic and it is not solely used in developing countries, but lately it is trend observed across a number of countries. According to Kang and Miller, in USA around 45% of the value of federal procurement contracts arise from tenders featuring single-bid [7]. Unfortunately, similar trend is observed in the EU, where according to The European Court of Auditors there is a significant rise in the share of single bidding from 23.5% in 2011 to 41.8% in 2021 [8].

Bulgaria is not excluded from this negative trend in public procurement and over the past years, despite the changes in legislation, there is a slight increase in single-bidding (see Fig. 1). Considering that there are cases when due to natural monopolies or patented products, single-bidding is inevitable, the European Commission in their Single market and competitiveness scoreboard has accepted that single-bidding below 10% is acceptable, while anything above 20% is unsatisfactory and indicates high risk of corruption. Therefore, such high scores regarding the public procurement in Bulgaria are indicative of a persisting problems that can lead to corruption.

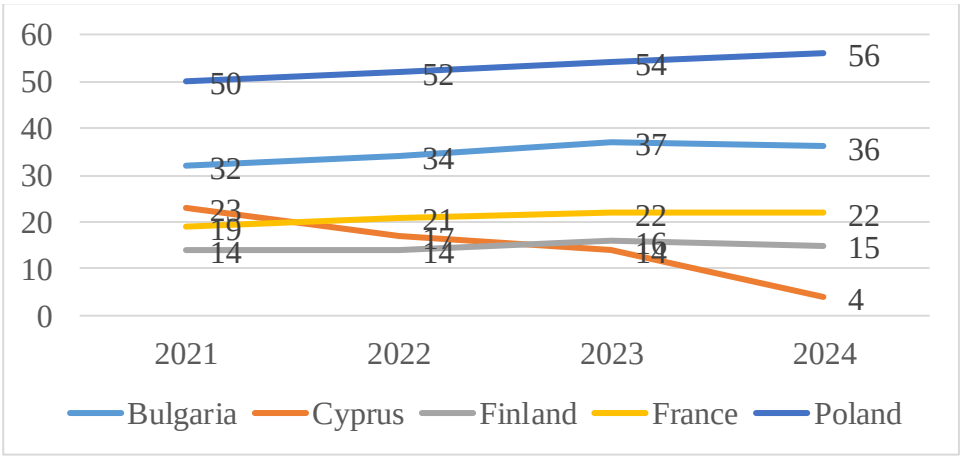


Fig. 1. Single-bidding in public procurement, 2021-2024 [9]

Bearing in mind, some of the single-bid contracts in Bulgaria are inevitable due to natural monopoly, there are a number of other plausible causes that contribute to this and

therefore increase the risk of corruption. The first and rather obvious reason is the customized (tailored) tender specifications, in other words the contracts' specifications that in a way can deter most companies to participate in a given public procurement. A number of studies have identified that the Bulgarian procurement system has a high share of tailored specifications, usually characterized by imposing very technical requirement that rarely can be met by more than one bidder [10]. Despite, it is a well-known corruption practice, there are no accessible data to indicate to what extend the single-bidding in Bulgaria is due to this fact.

There are however a number of other plausible causes that can be measured and respectively contribute to the high numbers of single-bidding in Bulgaria. One of these is the lack of publicity or insufficient information regarding public procurement that can lead to low participation [11], respectively increasing the corruption risk by creating conditions for the participations predominantly on firm in close relations to the government. Considering the existing information about the number of procurement contracts signed contracts without competition through a negotiated procedure without prior publication in Bulgaria, it can be concluded there is a significant corruption risk.

Another quite common practice in public procurement is the use of the so called restricted procedures. In some cased this practice does not pose risk of corruption, on the contrary its use is credible for example due to national security risk. However, the overuse can indicate higher risk of corruption. Following the EU regulation, Bulgaria is obliged to allow anyone asking to participate in such procedure, but only the pre-selected bidder can submit tenders [12], thus restricting the number of final bidders to the "chosen" ones.

The available data in Bulgaria on these types of procurements shows that this practice is overused and the leading one of them is the so called in-house contracting. For example, according to the Bulgarian Minister of Finance, the share of in-house contracting is approximately 42% of all government contracts, estimated to 4.4 billion euro [13].

Despite the fact that this practice is not illegal in Bulgaria, its vast application in especially in the construction and infrastructure sector rises a lot of questions, because the winning bidders often does not have the sufficient resources (human and equipment) to fulfil the contract alone. Therefore, they subcontract private companies for certain or all activities, evading the procurement procedures. This creates rather favourable conditions for corruption.

The political tensions in the last years and the increased sensitivity of the civil society on this matter have led to a significant restriction of the use of in-house procedures in Bulgaria in 2022, thus aiming to curb the direct, non-competitive awarding of contracts by prohibiting subcontracting. Despite this, the available data for 2024 is not flattering for Bulgaria. Compared to all the other EU countries, the proportion of procurement procedures negotiated with a company without any calls for bid is highest one in the union – 20% of all procurement, while it is considered appropriate up to 5% [9]. Therefore, we can conclude that Bulgaria is among the countries that face higher corruption risk in relation to this indicator and despite the gradual decline in the use of such practices over the past few years (see Fig. 2), it remains rather high and requires additional measures to tackle the problem.

Some sectors in Bulgaria are more prone to single-bidding contracts than others, especially prevalent in sectors with high-value national projects. The answer why is a combination of factors, but mainly two of them have sugnificant importance – politically connected firms and not that strict oversight, compared to project, funded by EU funds. Based on the information, derived from the European Court of Auditors, Kr. Valcheva oulines that the biggest share of single bidder contracts in Bulgaria for the period 2011-2021 is observed in the sector of transport services – 71.2%, followed closely by the energy sector – 63.7% and the sewage sector – 60.3%. The lowest share is observed in the sector of medical equipment – 23.6% and financial services – 30.1% [14]. Evidently, even the sectors where we observe the lowest share

of single-bidding, it reamins above the EU average and it is a clear indication either for lack competitiveness or "state capture".

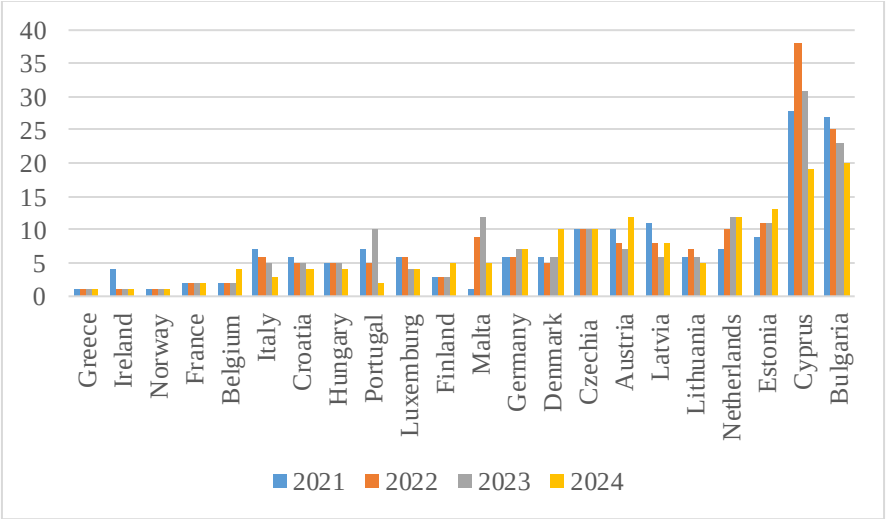


Fig. 2. Trend in the use of direc awards in EU – 2021-2024, % [9]

Lastly, the bigger share of single-bidding in public procurement in Bulgaria is also fueled by the rather high number of contracts awarded without competion using the so called direct negotiation. Although, the Bulgarian legislation stipulates certain limitations of using this tool, the trend from the last years negative – in 2022 approximatelly 4.55% of all contracts are signed by direct negotiations, in 2023 they rise to 4.94%, while in 2024 the data shows increase to 6.54% [2]. This increase contributes to the overall high share of direct awards in Bulgaria and can be a sign for a failed competition process, lack of tranparency and deeply rooted connections between the political elite and some of the participating companies in public procurement.

4. Political favouritism in public procurement in Bulgaria

Fair competition in public procurement in Bulgaria in significantly influenced by political connections. According to Fazekas, Ferralli, and Wachs, political favouritism in public procurement refers to the allocation and performance of government contracts by violating the principles of open and fair procurement in order to favour connected bidder(s) to the detriment of all others [6].

The most common indications for political favouritism in public procurement are single-bidding and tailored tender specifications, rather common in Bulgaria case. Political favouritism is extremely damaging not only for the public procurement, but for the economic competitiveness and productivity. However, it is rather profitable for some companies, because they tend to have higher corporate profits, due to the preferential access they have to uncompetitive tenders.

In 2025, the World Bank in a study "Procuring growth. The impact of political favouritism on public procurement and form performance in Bulgaria" raised awareness about the gradual increase in the percentage of politically connected firms that win bids in public procurement in Bulgaria. It has been estimated that they represent 0,9% of all firms, however

the average size of public procurement contracts awarded to politically connected firms is approximately 1.13 million euro, which is almost double compared to the average size of contracts awarded to unconnected firms [6].

Although it is not that easy to measure the percentage of political favouritism in public procurement, there are a number of red flags that indicate a significant risk factor for that. There is no need to explain further that one of the most common risk factor for favouritism is the high frequency of single-bidding, which is some of the cases conceal favouritism.

Red flags for political favouritism are also evident in the abuse of power of public official that due to their position can influence in a way the procurement process. This can be done by imposing tailored technical specifications – a case visible in 2021 in a public procurement by the Ministry of defence for the purchase of automobiles, in which the technical specifications are formulated in a way that only one specific model can be delivered [15].

Another example for abuse of power is the exercised pressure by political figures on members of the procurement commission, in order to disqualify some of the candidates and to award the contract to a politically connected firm.

In some cases, political favouritism is evident from the short submission periods and/or rapid decision period to award a contract.

Although the EU and Bulgaria legislations allow in emergency periods to implement changes regarding the public procurement process, political favouritism can be clearly noted by the misuse of the so called "emergency" exemptions, in order to avoid competitive procedures, thus a politically close firm can benefit from a direct award. The misuse of this exemption in Bulgaria to some extent is possible because of the rather broad definition of the term "extreme circumstances" in the Public Procurement Act.

In order to curtail this misuse and reduce the percentage of political favouritism due to misuse of emergency exemptions, in 2023 the Public procurement agency updated its Methodological Instruction MU No. 2/21.11.2022 on the award of public procurement contracts with negotiated procedures without prior publication, on the basis of Art. 79, para. 1, item 4 of the Public Procurement Act, by narrowing the scope of emergency by excluding events and/or facts that hinder/disturb the normal performance of the contracting authority's legally established activities from the definition.

The last red flag that often is seen as an indicator for political favouritism are the repeated wins by specific firms – rather obvious pattern, where some well-connected firms consistently win large-scale projects, especially in construction and energy win bids constantly [16], such as Bulgarstransgaz EAD, Glavbolgarstroy, Hydrostroy, Patinzheneringstroy-T EAD, Trace Group Hold – PLC, etc.

5. Conclusion

Since 2020, the European Commission and the European Court of Auditors observe a rather worrying trend in the EU, namely rise in the single-bidding in public procurement, which is some countries in addition is accompanied by data for relatively high share of political favouritism. Considering all the available data, we can conclude that Bulgaria is among the countries in the EU that have one of the highest percentage of single-bidding in public procurement, as well as there are evidences for close political connections between political officials and some of the major bidders – key red flags for corruption risk.

In order to tackle this problem, Bulgaria undertakes a number of legislative amendments, however the outcome continues to be not very favourable. Although the data shows slight decrease in the percentage of direct awards in public procurement, the numbers remain one of the highest in the EU and all reports from the European Commission and the World Bank state that this hinders the economic competitiveness and therefore the state will

benefit greatly by implementing more open public procurement processes, thus stimulating open and competitive market and reducing the costs for public services.

References

1. European Commission: Commission Staff Working Document. Evaluation on Directive2014/23/EU on Concessions, Directive 2014/24/EU on Public Procurement and Directive 2014/25/EU on Utilities, Brussels, 14.10.2025 (https://ec.europa.eu/transparency/documents-register/detail?ref=SWD(2025)332&lang=en)
2. Годишен доклад на Агенцията по обществени поръчки, 2024 (https://www2.aop.bg/wp-content/uploads/2025/04/GD_AOP_2024.pdf)
3. Rose-Ackerman, Susan, Palifka, Bonnie. Corruption and Government: Causes, Consequences, and Reform. Cambridge: CUP, 2016.
4. Miteva, Mihaela. Bulgaria probes Lukoil's local units for competition violations, 13.06.2025 (https://seenews.com/news/bulgaria-probes-lukoils-local-units-for-competition-violations-1276905)
5. European Commission: Commission Staff Working Document: 2025 Country Report – Bulgaria, 04.05.2025 (https://kmp.respaweb.eu/download/doc/2025%20Country%20report%20-%20Bulgaria.pdf/8e21b44fe2cc5aa7ff607a045934d44c.pdf)
6. World Bank Group: Bulgaria. Public Finance Review, 2023 (https://documents1.worldbank.org/curated/en/099122123051567091/pdf/P18022318324dc060196411826bccae6a6e.pdf)
7. Kang, Karam, Miller, Robert. Winning by Default: Why is There So Little Competition in Government Procurement? // The Review of Economic Studies. Vol. 89 (3), 2021, pp. 1495-1556.
8. The European Court of Auditors: Public procurement in the EU: Less competition for contracts awarded for works, goods and services in the 10 years up to 2021, 2023 (https://www.eca.europa.eu/ECAPublications/SR-2023-28/SR-2023-28_EN.pdf)
9. European Commission: Single market and competitiveness scoreboard, 2024 (https://single-market-scoreboard.ec.europa.eu/business-framework-conditions/public-procurement_en)
10. Stoyanov, Alexander, Gerganov, Alexander, Yalamov, Todor. State Capture Assessment Diagnostics. Sofia: CSD, 2019.
11. Coviello, Decio, Mariniello, Mario. Publicity requirements in public procurement: Evidence from a regression discontinuity design // Journal of Public Economics. Vol. 109, 2014, pp. 76-100.
12. SELDI: Public procurement irregularities – hidden risks for the post-Covid economic recovery, 2021 (https://seldi.net/public-procurement-integrity-in-southeast-europe/public-procurement/#r61)
13. Соколова, Цветелина. Служебният кабинет спира харченето на милиарди без обществени поръчки // Mediapool.bg, 30.06.2021 (https://www.mediapool.bg/sluzhebniyat-kabinet-spira-harcheneto-na-miliardi-bez-obshtestveni-porachki-news323641.html)
14. Valtcheva, Krasimira. Public procurement in Bulgaria: Do we get value for the taxpayer's money // Journal of management Sciences and Application. No I, 2024, pp. 42-52.
15. Атанасова, Мария. Антикорупционният фонд: Министерството на отбраната си е поръчало Аудита А8 за 314 000 лева, 02.06.2021 г. (https://fakti.bg/bulgaria/584274-antikorupcionniat-fond-ministerstvoto-na-otbranata-si-e-porachalo-audita-a8-za-314-000-leva#google_vignette)
16. Mungiu-Pippidi (ed.) Government Favouritism in Europe. The Anticoruption Report 3. Barbara Budrich Publishers, 2015 (https://csd.eu/fileadmin/user_upload/publications_library/files/2015/13_00_All_ACRVolume3_Crime.pdf)

Selection of an Alternative for Reducing the Risk of Forest Fires in the Black Sea Region through the Application of a Payoff Matrix and Simulation Modeling

Anna-Maria Markova, Yasen Tsvetkov

University of National and World Economy, Sofia, Bulgaria
annikolaeva@e-dnrs.org, qsen.cvetkov@abv.bg

Abstract. The Black Sea region is characterized by multiple features that influence the risks of forest fires. This necessitates the search for new, more advanced methods for assessing investment alternatives and rational selection of an alternative for investment in actions to reduce forest fire risk. The paper proposes a model that combines the well-known utility matrix model and its further development through the use of simulation modeling. Three alternatives for building observation towers are evaluated. Based on hypothetical information, a deterministic utility matrix model is initially developed. The model is then converted into a simulation model using specialized software. A total of 10,000 simulations are performed, the results of which are analyzed using statistical methods. Based on the results, a rational choice of alternative is made.

Keywords. Forest Fires, Black Sea Region, Utility Matrix, Simulation Modeling

1. Introduction

This paper examines the characteristics of the Black Sea region in Bulgaria from the perspective of the specific nature of forest fires occurring there. The aim of the study is to propose an alternative method for evaluating and selecting an alternative for investment in reducing forest fire risk in the Black Sea region. The increasing complexity of forest fire risk in contemporary conditions requires the application of various actions tailored to the specifics of the region for its management [1, 2]. The introduction of new technologies for forest monitoring supports this process. The question arises as to which of these technologies should be applied and with which specific projects they should be implemented in practice so that public resources allocated to them are used in a rational manner.

For this purpose, more reliable methods are needed to evaluate investment alternatives, taking into account expected costs, expected damages under different scenarios, and the presence of a high degree of uncertainty in the decision-making environment. In theory and practice, some reliable decision-support tools have long been used in such complex situations, such as the utility matrix. On the other hand, methods that support the evaluation of alternatives in highly volatile environments, such as simulation modeling, have received significant development. Moreover, contemporary software products supporting the development and application of simulation models are so advanced that they can be used in complex and rapidly changing environments with relatively modest resources.

2. Problem Identification

The Black Sea region of Bulgaria is characterized by multiple features that determine the specificity of forest fire occurrence and development. On the one hand, certain parts of the

region (e.g., Dobrudzha) are predominantly flat or gently hilly, which allows for the occurrence of strong winds. Agriculture is highly developed in the area, characterized by intensive use of high-performance agricultural machinery. At the same time, drought is a frequent phenomenon. This, combined with frequently observed high temperatures, increases the probability of forest fire occurrence and rapid spread. Other parts of the region (e.g., Strandzha, in the Primorsko – Rezovo section) consist of mountainous, difficult-to-access terrain with dense broadleaf forests, which complicates access for heavy firefighting equipment. The entire Black Sea region is strongly touristic. Simultaneously, there are dense forests near resort and beach areas. Tourists do not always comply with fire safety rules. Illegal camping in prohibited areas is common. These factors also increase the likelihood of forest fire occurrence. The causes of forest fires in the region are diverse and can be divided into three groups: human-induced (anthropogenic), natural factors, and combined. Examples include sparks from poorly maintained or operated agricultural machinery, power line accidents, improper fire lighting by tourists in unsafe locations, discarded unextinguished cigarettes, violations of fire safety rules during hot work (welding, metal cutting, using angle grinders), and the burning of stubble and dry grass areas. Natural causes include lightning igniting dry vegetation or trees, spontaneous combustion of humus or peat, and improperly stored waste. Forest fires often arise from a combination of these two groups of factors.

Once a forest fire occurs, its development and speed depend on existing conditions. Literature reports that fire behavior is determined by the so-called fire behavior triangle, which includes three factors: fuel, weather, and topography [3, 4]. Various alternatives exist to reduce forest fire risk in the region, differing in factors such as required implementation costs and expected benefits. International practice presents diverse methods for comparative analysis and alternative selection in the field of security, but in all cases, there are difficulties in collecting the necessary information.

3. Methodology of the Study

This paper proposes a method for selecting an alternative to reduce forest fire risk, based on the application of the payoff matrix method in combination with simulation. A payoff matrix is constructed, with rows representing alternatives for reducing fire risk that need to be identified. These may include clearing firebreaks, building forest roads, acquiring new and upgrading existing firefighting equipment, among others. For the purposes of this paper, the evaluated alternatives involve building observation towers using different technologies. Columns of the matrix represent natural conditions, i.e., possible scenarios for forest fire occurrence and development. Cells of the matrix contain expected utilities for each combination of selected alternative and natural state.

The utility matrix is a widely used method for justifying rational decisions in environments of risk and uncertainty. Depending on whether the selected utility is evaluated positively (e.g., profit, revenue) or negatively (e.g., costs, losses), the matrix is used to search for either the maximum or the minimum utility.

In turn, there are various decision criteria that reflect the decision-maker's attitude toward risk. Highly optimistic decision-makers are advised to apply the maximax criterion. The choice of an alternative under this criterion is made by first identifying the maximum utility for each alternative, and then selecting the alternative that provides the highest of these maximum utilities.

Highly pessimistic decision-makers apply the maximin criterion. Under this criterion, the minimum utility for each alternative is first determined, and then the alternative with the highest of these minimum utilities is selected.

In cases where the decision-maker wishes to consider the expected utilities across all scenarios, the Laplace criterion may be applied. This is possible only when the probabilities of occurrence of each scenario are approximately equal. To make a decision using this criterion, the average utility is calculated as the arithmetic mean of all utilities for each alternative. The alternative with the highest average utility is then selected.

When the selected utility is negatively evaluated, it is recommended that optimistic decision-makers apply the minimin criterion, while pessimistic decision-makers apply the minimax criterion. The Laplace criterion may also be applied in this case, but the chosen alternative will be the one that provides the minimum average utility.

In certain situations, the decision-maker is interested in selecting an alternative that minimizes opportunity loss. In such cases, the minimax regret criterion is recommended. There are also many other criteria that can more precisely reflect the decision-maker's attitude toward risk. Additionally, some criteria can be applied when information about the probabilities of each scenario is available. In such cases, it is possible to apply the lowest expected utility criterion.

In this study, utility is measured as the sum of investment costs for constructing the facility and expected damages under each scenario [5]. Three hypothetical options for building fire observation towers are defined, with three main scenarios of fire occurrence depending on local climatic conditions. The first option involves constructing simple wooden towers for manual observation at varying heights, without technical means, which are the cheapest and have limited observation range. The second option is the acquisition of portable observation stations mountable on vehicles with limited automation.

They can be deployed in different locations according to risk and do not require year-round maintenance, but their deployment requires transportation and road infrastructure, and placement on high points with good visibility may be challenging. The most expensive alternative is constructing stationary towers with modern equipment, including antennas, cameras, and radio communication. Data is processed using state-of-the-art computing technology. They have the widest observation range and are most suitable for early fire detection.

Scenario one assumes low fire danger, with low temperatures, high fuel humidity, and weak winds. Scenario two assumes dry conditions, low temperatures, and weak winds. The highest risk scenario includes drought, high temperatures, and strong winds. For each natural state and each equipment, purchase alternative, utility values (Rij) for investment costs and potential damages are defined. Minima of the compared utilities are sought. Probabilities of each scenario occurring in the Black Sea region are hypothetically assigned using expert methods. Based on these considerations, a deterministic utility matrix model is developed.

It is then assumed that, due to the specifics of the Black Sea region, the selected utilities can be treated as random variables. Expert evaluation defines these variables as uniformly distributed within a known interval. This results in a probabilistic model subjected to simulation analysis [6, 7].

The software @Risk by Lumivero is used, with 10 000 simulation runs performed and results statistically analyzed.

4. Model Development and Simulation

For this study, a hypothetical situation is developed, and a deterministic utility matrix model is compiled (Table 1). Alternatives are evaluated according to common criteria: minimax, minimax regret, Laplace, and expected utility.

Table 1. Utility matrix for selecting an alternative to reduce forest fire risk in the Black Sea region, thousand €

Alternatives	High Danger	Medium Danger	Low Danger	Minimax	Maiximin	Mean Utility
1	900	520	3	3	900	474.3
2	500	470	450	450	500	473.3
3	5	400	600	5	600	335
Min utility	5	400	3			

Under the minimax criterion, the alternative with the lowest value under the most favorable condition and lowest utility level is selected – Alternative 1 for the simple structure. Under the minimax criterion assuming high utility scenario, Alternative 2 is chosen. For minimax regret, a matrix is constructed by calculating the differences of each cell from the lowest utility values of Alternatives under each natural state (Table 2). The alternative with the lowest maximum regret, Alternative 2, is selected.

Table 2. Alternative loss matrix from Table 1, thousand €

Alternatives	High Danger	Medium Danger	Low danger	Max regret
1	895	120	0	895
2	495	70	447	495
3	0	0	597	597

Under the Laplace criterion for mean utilities, Alternative 3 with 335 thousand € mean utility is advisable. Considering uncertainty, the expected utility criterion is applied with assigned probabilities: low danger p1=0.2, medium p2=0.5, high p3=0.3. The expected utilities for each alternative are: 440.9, 470, 381 thousand €. Calculations can be performed using Excel [8, 9]. The model is further developed using @Risk software [10, 11]. Input variables are the defined utilities with expert-assigned uniform distributions, and output variables are expected utilities for each alternative and differences between alternatives. The model is simulated 10,000 times. Results of statistical analysis of Expected Utility for three alternatives are shown in Table 3.

Table 3. Expected Utility for three alternatives

Alternative	Min	Max	Mean	Mode	Median	STD.Dev
1	482.323	575.741	530.6	531.634	530.661	15.281
2	435.65	514.65	475	477.941	475.037	11.979
3	291.561	349.561	321.5	319.219	321.446	9.589

The statistical distribution of Expected Utility results is shown on Figure 1. The model identifies the alternative providing minimum utility, which is Alternative 3 with the lowest Mean Expected Utility. Alternative 2 ranks second, and Alternative 1 third. Assuming standard deviation of Expected Utility as a risk measure, Alternative 3 has the lowest risk.

A comparison between Alternative 1 and 2 shows a probability of reversal.

Alternative 1 may have better Expected Utility than Alternative 2. The software allows calculation of this probability, useful when ranking alternatives by utility.

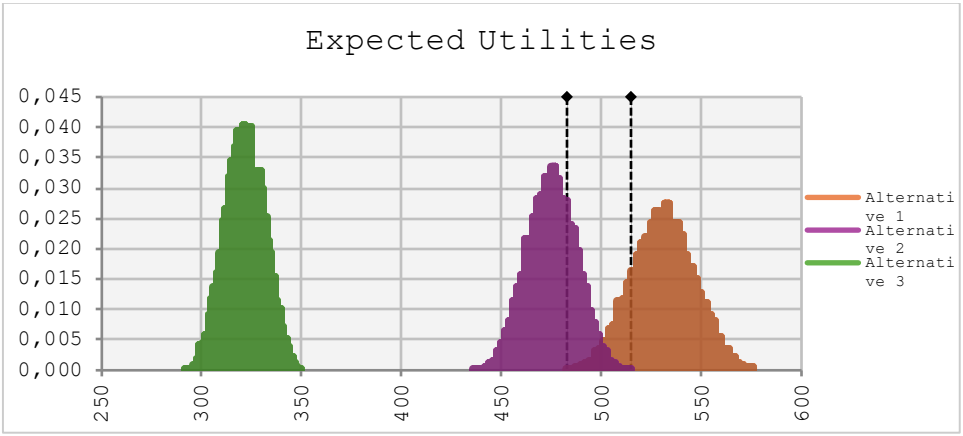


Fig. 1. The statistical distribution of Expected Utility results

A frequently observed phenomenon in practice is the comparison of two alternatives based on the expected utility criterion and the level of risk. From a research perspective, a particularly interesting case arises when one alternative provides a higher (or lower, when seeking the lowest expected utility) expected utility, while at the same time being characterized by a higher level of risk. In such a situation, it is not possible to directly determine the better alternative. In these cases, the "probability of reversal" method can be applied. In other words, the probability is calculated that, due to the dynamics of utility values, the alternative with the higher expected utility may actually result in a lower realized utility. The recommendation in this case is that the decision-makers should choose an alternative by taking into account both the probability of reversal and their own attitude toward risk.

5. Conclusion

As a result of the study conducted, it can be concluded that the Black Sea region presents a complex environment with a high degree of uncertainty regarding forest fire risks. Simultaneously, contemporary tools and software products are available that can be effectively applied to justify decisions in selecting alternatives for investment projects aimed at reducing these risks. Additionally, the necessity of collecting and synthesizing expert opinions, which serve as the foundation for model development and decision justification, must be recognized. A critical long-term objective should be the establishment of a system for gathering, consolidating, and systematizing information on forest fires occurring, their causes, scenarios of development, and the methods used for their suppression.

The application of modern methods for selecting an alternative, such as the utility matrix, in combination with Monte Carlo simulation methods, enables the aggregation of expert opinions and the rationalization of the decision-making process. At the same time, it allows for a precise assessment of the decision-maker's attitude toward risk. This is particularly important in high-risk situations, such as forest fires in the Black Sea region.

Acknowledgements

This work has been carried out in the framework of the Project "Smart Risk Management for Business from Adverse Events and Natural Disasters", supported by the National Science Fund of Bulgaria, Grand № KP-06-H55/5 from 15.11.2021.

References

1. Tedim, F., Leone, V., & McGee, T. K. (Eds.). (2020). Extreme Wildfire Events and Disasters: Root Causes and New Management Strategies. Elsevier.
2. Tsvetkov, T., & Tsvetkov, Y. (2023). Possibilities of Applying the Bibliometric Analysis Methods of Scientific Publications on the Relationship Between Climate Change and the Number and Intensity of Forest Fires. 13th International Conference on Application of Information and Communication Technology and Statistics in Economy and Education (ICAICTSEE – 2023), December 15-16, 2023, UNWE, Sofia, Bulgaria. ISSN 2367-7643 (Online).
3. Finney, M. A., McAllister, S. S., Grumstrup, T. P., & Forthofer, J. M. (2021). Wildland Fire Behaviour: Dynamics, Principles and Processes. CSIRO.
4. Meldrum, J.R. (Ed.) (2022) Wildfire Hazard and Risk Assessment. MDPI. ISBN 978-3-0365-3581-4
5. Chandra, P. (2019). Projects: Planning, Analysis, Selection, Financing, Implementation, and Review (9th ed.). McGraw Hill Education (India).
6. Arnold, T., & North, D. (2011). Risk Simulation Concepts and Methods. In H. K. Baker & P. English (Eds.), Capital Budgeting Valuation: Financial Analysis for Today's Investment Projects (pp. 279-296). John Wiley & Sons.
7. Verschuuren, G. M. (2027). 100 Excel Simulations: Using Excel to Model Risk, Investments, Genetics, Growth, Gambling and Monte Carlo Analysis. Holy Macro.
8. Abdul Karim, S. A., & Shafie, A. (Eds.). (2022). Towards Intelligent Systems: Modeling and Simulation with Applications to Energy, Epidemiology and Risk Assessment. Springer.
9. Guerrero, H. (2019). Excel Data Analysis, Modeling and Simulation (2nd ed.). Springer.
10. Brandimarte, P. (2014). Handbook in Monte Carlo Simulation: Applications in Financial Engineering, Risk Management, and Economics. Wiley.
11. Shonkwiler, R. W., & Mendivil, F. (2009). Explorations in Monte Carlo Methods (2nd ed.). Springer.

Technological Innovations and Cooperative Framework in Nuclear Security: Strengthening Integrated Border Security System for Global Stability

Sara Khalid Khider Mohamed

University of National and World Economy, Sofia, Bulgaria
sarakideir@gmail.com

Abstract. The purpose of this paper is to examine how modern technology is transforming border security in response to transnational threats and nuclear or radiological material trafficking. An analysis of the study examines how artificial intelligence (AI), biometric identification, blockchain, and unmanned systems are transforming borders from static physical barriers to intelligent and adaptive systems. Additionally, it analyzes how these technologies align with international frameworks led by the International Atomic Energy Agency ,the primary focus is on strengthening nuclear security by preventingthe illicit movement of dangerous materials. Furthermore, it aims to improve the physical and cyber resilience of borders. A key finding of the paper is that cross-border threats must be detected and responded to more effectively through cooperation and data sharing among nations. As a result, the future of border security hinges on the integration of technology, governance, and international collaboration. By doing so, secure, efficient, and globally coordinated systems will be created.

Keywords: Border security; nuclear security; artificial intelligence; unmanned aerial vehicles; international cooperation; global stability.

1. Introduction

Due to the globalization of trade, technology, and mobility, the goal of ensuring a secure national border has become crucial to global stability. The management of borders is no longer just a matter of physical demarcation; it involves an integrated system of surveillance, intelligence, and governance designed to protect sovereignty while facilitating lawful movement and trade. Increasing threats, like terrorism and cybercrime, along with illicit traffic of nuclear and radiological materials, place border security at the intersection of national defense, international cooperation, and technological innovation (Andreas, Snyder, 2021).

In the 21st century, transnational activities that exploit weak borders have risen dramatically, including smuggling weapons, human trafficking, and diversion of nuclear materials. More than 150 unauthorized movements of nuclear or radiological materials were reported globally between 2020 and 2023, according to the International Atomic Energy Agency (IAEA, 2024). Therefore, robust border architecture that can detect, verify, and respond successfully in an early stage are urgently needed. The UN Security Council Resolution 1540 (UNSCR 1540) codifies the global obligation of states to prevent non-state actors from acquiring or transporting weapons of mass destruction, relating border management directly to international nuclear security.

As a result, border control serves as the first line of defense against the illicit transfer of nuclear materials, radiological sources, and dual-use technologies. The integration involves national agencies, regional organizations, and global frameworks such as INTERPOL, IAEA, and World Customs Organization (WCO).

Despite this, this transformation is not easy. A balance must be struck between sovereignty and privacy, as well as transparency and international cooperation. Cyberattacks on customs databases and manipulation of radiation detection systems are becoming more common as border systems become more interconnected. Achieving border security resilience requires interoperability, cyber protection, and legal and ethical frameworks for sharing data and deploying technology (UNODC, 2022).

As nuclear security encroaches on global stability, it is vital to integrate border security systems. It examines how emerging technologies, international cooperation, and policy innovation collectively shape the global response to cross-border threats. Border security is analyzed in the following sections, the technological foundations are assessed, cooperation frameworks are explored, and pathways for strengthening resilience are proposed.

2. Theoretical and conceptual background

2.1. Globalization and Border Security

In the past, borders were used to define national sovereignty as static physical barriers, such as walls, checkpoints, and natural barriers. The modern border has evolved into a dynamic, multi-layered security environment integrated with real-time surveillance, biometric identification, and data-driven intelligence (Garcia, 2023). This shift aligns with the concept of smart borders, where technology facilitates as well as controls border crossings. Through automation and coordination among agencies, efficiency in movement and security in verification are achieved.

A theory of border security that draws on systems theory and international security studies underpins modern border security. As a result of shared data, harmonized standards, and collective enforcement mechanisms, borders now function as nodes within a global governance network (Anderson, 2021). By preventing illicit transnational flows that could destabilize regions or empower non-state actors, border security contributes not only to national defense but also to international conflict prevention.

2.2. Border Control as a Nuclear Security Component

Detecting, deterring, and interdicting the illicit trafficking of nuclear and other radioactive materials is a critical function of border control within the nuclear security domain. According to IAEA Nuclear Security Series No. 15, detection architectures must extend across borders, ports, and airports. A radiation portal monitor, a mobile detector, and an analytical command center will be connected through a secure communication network in such an architecture (IAEA, 2024).

Moreover, international cooperation mechanisms such as the Global Initiative to Combat Nuclear Terrorism (GICNT) and the Proliferation Security Initiative (PSI) assist member states in strengthening their border capabilities. By implementing these frameworks, customs authorities, intelligence agencies, and law enforcement can exchange information in real-time about nuclear material trafficking incidents (INTERPOL, 2022)

2.3. Technology and policy integration

The effectiveness of border security is increasingly dependent on the fusion of technology and governance. As a result of artificial intelligence (AI), predictive analytics can detect anomalies in cargo and passenger data, while blockchain can ensure the integrity of customs documentation. By integrating biometric databases and automated control gates

across Schengen Area borders, the European Smart Borders Project (European Commission, 2024) demonstrates how digital transformation enhances security and mobility.

From a policy standpoint, these innovations necessitate harmonized regulatory frameworks, standardized data protection measures, and ethical oversight to ensure transparency. The transition toward integrated systems, therefore, represents not only a technological revolution but also a transformation in institutional design one that balances efficiency, accountability, and human rights.

3. The Role of Border Control in Global Security

3.1 Strategic Importance of Border Management

The control of borders plays a crucial role in maintaining national sovereignty and international peace. They regulate the movement of people, goods, and vehicles, forming the first layer of defense against illicit activities such as terrorism, arms trafficking, and nuclear material smuggling (ICMPD, 2022). Since the end of the Cold War, border control has evolved from static enforcement to an intelligence-based, dynamic paradigm. As a result of the integration of surveillance, data analytics, and international cooperation, modern systems can address threats that transcend physical boundaries (Miller & Johnson, 2023).

Border management underpins global trade, migration, and crisis response. According to the World Customs Organization (WCO, 2024), integrated border management (IBM) models where customs, immigration, and law enforcement collaborate under shared data platforms have reduced cross-border security breaches by over 30% in regions adopting full interoperability. This operational shift has also improved response time to transnational incidents, enabling faster identification of high-risk cargo and individuals through pre-screening systems such as Advanced Passenger Information Systems (APIS) and Passenger Name Record (PNR) databases (Gonzalez & Lee, 2024).

3.2 Enhancing border security to prevent illegal trade and terrorism

A border control system serves as both a preventive and a reactive mechanism against transnational threats. In absence of effective border governance, smugglers may be able to move illicit goods, weapons, and radiological materials undetected across borders. Organized crime is largely enabled by border corruption, inadequate technology, and fragmented coordination, according to the United Nations Office on Drugs and Crime (UNODC, 2023) Furthermore, the IAEA's Illicit Trafficking Database (ITDB) records numerous instances where poorly secured checkpoints were exploited to move radioactive materials.

There is also a persistent threat of terrorism. The movement of explosives and CBRN (Chemical, Biological, Radiological, and Nuclear) materials across conflict zones and fragile states is facilitated by porous borders. Accordingly, INTERPOL's Global Border Management Programme (2024) emphasizes harmonizing CBRN detection and training among its member states. As part of this initiative, border enforcement and intelligence-sharing tools are integrated, including biometric alerts and risk profiling tools.

3.3 Economic and Political Dimensions

Border systems play a crucial role in economic resilience beyond security. By preventing counterfeit goods and ensuring customs compliance, they facilitate legitimate trade. A Schengen Information System (SIS II) which enables real-time exchanges of data on goods and individuals across borders is a crucial component of border interoperability in regions such as the European Union (EU) (Garcia, 2022). Several geopolitical pressures - including

migration crises, regional conflicts, and sanctions regimes - often test the adaptability of these systems. As is seen in Eastern Europe and North Africa, border management has become a geostrategic tool, influencing diplomatic relations and power dynamics (Anderson, 2021).

4. Cross-border and nuclear threats

4.1 Transnational Security Threats

Modern security threats transcend national borders and require networked responses. In addition to terrorism, cybercrime, drug trafficking, human smuggling, and the movement of dual-use materials, there are other threats. By interconnecting physical and digital domains through electronic documentation, blockchain logistics, and AI-enabled data exchange, hybrid vulnerabilities have arisen. A cyber-attack on customs or radiation detection databases could compromise entire regions' ability to intercept dangerous materials (UNODC, 2023)

In addition, technological democratization has allowed non-state actors access to advanced tools such as drones and encryption software, enabling covert transport and communication. Therefore, border management must integrate cyber-border security, protecting both physical and digital borders.

4.2 Nuclear and Radiological Material Trafficking

In terms of international peace and stability, the trafficking of nuclear and radiological materials is one of the most pressing challenges. During the period 2020-2023, the IAEA (2024) confirmed more than 150 incidents of radioactive substances in unauthorized possession or movement. The incidents cannot all be malicious, but each of them can lead to nuclear terrorism or contamination of the environment.

Detection architecture at border crossings and maritime ports integrate radiation portal monitors, gamma spectrometers, and neutron detectors to mitigate this risk. A growing number of these systems are supported by AI algorithms that analyze spectral data in real-time to distinguish benign sources from suspicious ones (Patel &Nguyen, 2024). In addition, mobile detection units are deployed in remote areas or during emergency response operations, forming a multi-layered detection network.

4.3 Hybrid Threats and Dual-Use Technologies

Dual-use materials, substances and technologies that can be used for both civilian and military purposes present regulatory challenges. It is possible to divert high-grade materials into unauthorized hands because of weak export control systems and corruption at border posts. A Global Initiative to Combat Nuclear Terrorism (GICNT) emphasizes that no single state can address this risk in isolation; instead, international information exchange, shared intelligence, and coordinated interdiction exercises are essential (INTERPOL, 2024).

In addition to cyber threats, hybrid threats also involve physical threats. An adversary might use drones to map border surveillance zones or cyber intrusions to disable detection devices. Unmanned aerial systems (UAS) and AI-based recognition technologies are vital for reconnaissance and counter-surveillance (International Security and Technology Review, 2023)

4.4 Border Resilience and Risk Management

To achieve resilience, states must adopt a risk-based management approach that focuses on high-threat regions, dynamic intelligence assessment, and joint operational protocols. As an example, the Integrated Maritime Border Command of Australia uses satellite imaging to

detect illegal maritime entries (Taylor & Wong, 2021). Similarly, the European Union's *Frontex* agency applies AI-driven analytics to anticipate irregular migration flows, aligning operational planning with early warning systems.

In developing countries, capacity building is also essential for sustainable border operations. Due to limited funding, outdated equipment, and a shortage of skilled personnel, detection networks in Africa, the Middle East, and Southeast Asia are weak. Technical assistance has been provided by the International Atomic Energy Agency (IAEA) and the United States Department of Energy's Office of Nuclear Smuggling Detection and Deterrence (NSDD) to enhance these capabilities.

Table 1. Summary of Transnational Threats and Detection Technologies

Type of Threat	Description	Detection / Countermeasure Technology	International Framework Involved
Nuclear trafficking	Unauthorized transport of nuclear/radiological materials	Radiation portal monitors, neutron detectors, AI-assisted analysis	IAEA, GICNT, INTERPOL
Terrorism	Cross-border movement of weapons and operatives	Biometrics, PNR/APIS systems, AI profiling	UNSC 1540, INTERPOL
Cybercrime	Hacking of border and customs systems	Cyber firewalls, blockchain verification, quantum encryption	UNODC, WCO
Drug trafficking	Illicit movement of narcotics across land and sea	Cargo X-ray scanners, risk-based analytics	UNODC, WCO
Human trafficking	Smuggling of people across borders	Biometric identity systems, facial recognition	IOM, INTERPOL

[Compiled by the author based on IAEA (2024), INTERPOL (2024), and UNODC (2023)].

Figure 1 below illustrates an integrated border security model that incorporates nuclear detection, AI surveillance, biometric controls, and international coordination frameworks.

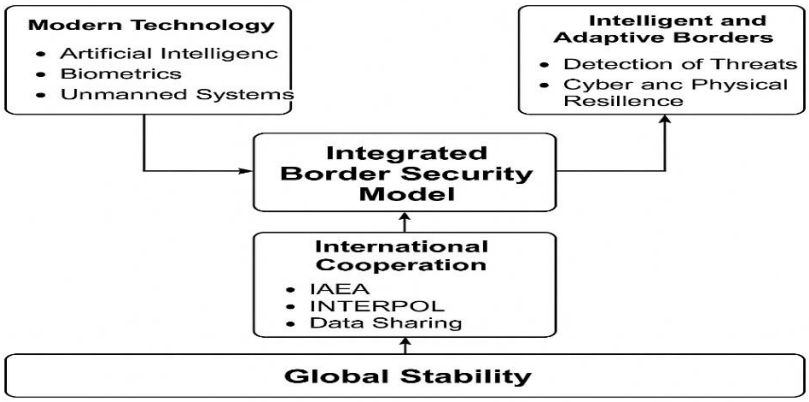


Fig. 1. Integrated border security model that incorporates nuclear detection, AI surveillance, biometric controls, and international coordination frameworks
[By the author from IAEA NSS-20-T (2024), INTERPOL border security frameworks.]

5. Components of Modern Border Security Systems

5.1 Detection and Control Multilayered Framework

A multilayered border security system integrates physical, digital, and organizational elements. These include biometric systems, automated border control kiosks, radiation detection portals, and intelligence-driven databases that share real-time information between agencies (WCO, 2022). By designing the system multilayered, a malfunctioning sensor doesn't compromise the entire system.

A key component of it is interoperability and data integrity, according to the International Centre for Migration Policy Development (ICMPD, 2024) A unified command and control center coordinates customs, police, immigration, and defense units as part of the Integrated Border Management (IBM) model. To reduce false positives and maximize manpower utilization, they use predictive algorithms to detect anomalies in cargo movement records or travel records.

5.2 Biometric Identification and Automated Systems

Identity verification at border checkpoints has been revolutionized by biometric technologies such as fingerprinting, facial recognition, and iris recognition. By reducing processing time and reducing human error, these systems maintain high levels of accuracy. Integrated biometric databases, linked to criminal watchlists and visa systems, enhance the identification of high-risk individuals, according to Puglise et al. (2022).

An automated border control system (ABC) combines biometric verification and self-service processing. In addition to improving efficiency the argue that contactless screening can also support pandemic-era protocols. Through the Smart Borders initiative, these tools are integrated into a single platform and demonstrated to be scalable across multinational zones.

At various international airports, automated border control (ABC) systems are used. Airports use ABC systems to accommodate the increase in air traffic and passengers, and to keep border crossings secure while promoting travel. ABC systems handle the instability caused by this development, such as disruptions at immigration and security departments or

delays in arrival patterns. In airports, face recognition, fingerprints, iris recognition, and E-passports are used. (Mamad Chaki, N., & Ashour, M. W. (2021).

5.3 Radiation Detection and CBRN Monitoring

To ensure nuclear security, radiation detection technology must be used. A growing number of border stations are equipped with gamma-ray imaging systems, mobile neutron detectors, and spectroscopic portal monitors capable of distinguishing natural radiation sources from artificial radiation sources. A central command unit can receive live data streams from detectors nationwide via digital integration, according to the IAEA (2024).

Moreover, handheld radionuclide identifiers (RIIDs) and vehicle-mounted spectrometers are used for mobile inspections. With AI-enhanced data processing, operators can interpret spectral anomalies more quickly, reducing the need for manual analysis. It is particularly useful in areas where illicit trafficking routes intersect porous land borders and maritime smuggling routes.

5.4 AI, Data Fusion, and Predictive Analytics

Almost all layers of border management are now supported by artificial intelligence (AI). To predict potential risks, machine learning models analyze massive datasets from customs declarations, shipping manifests, and travel records. In addition to reducing false alarms in cargo screening by over 40%, AI systems can improve radiological material detection sensitivity, according to Global Security Innovations (2022).

Data fusion platforms combine multiple data streams from satellite surveillance to biometric logs into a unified intelligence dashboard. The tool provides decision-makers with an understanding of risk clusters and allows them to deploy resources efficiently.

5.5 The Blockchain and Quantum Encryption

In cross-border operations, blockchain technology enhances data security and transparency. In addition to ensuring the authenticity of customs documents, it also prevents trade records from being manipulated (Lee, 2024). Each transaction is permanently logged, providing traceability for high-value or dual-use materials.

For border agency communications, quantum encryption is emerging as a next-generation safeguard. By encoding data at the quantum level, Quantum Key Distribution systems prevent interception or tampering, ensuring that even advanced cyberattacks cannot compromise critical border intelligence (European Commission, 2022).

6. International Cooperation and Policy Frameworks

6.1 The Imperative of Global Coordination

Nuclear and radiological threats are transnational, so no single state can maintain security alone. As a result of international cooperation, intelligence, best practices, and technologies are shared across jurisdictions. In establishing operational standards and facilitating joint exercises, organizations such as the International Atomic Energy Agency (IAEA), INTERPOL, the United Nations Office on Drugs and Crime (UNODC) and the World Customs Organization (WCO) play pivotal roles.

Over 100 countries are supported by INTERPOL's Global Border Management Programme (2024) deploying interoperable databases for criminal records and CBRN incident reports. To strengthen nuclear detection and interdiction capabilities, the Global Initiative to Combat Nuclear Terrorism (GICNT) encourages information exchange and tabletop exercises.

6.2 IAEA's Role in Nuclear Security Capacity Building

In addition to training workshops, equipment deployment, and Integrated Nuclear Security Support Plans (INSSPs), the IAEA coordinates a wide range of assistance programs. Member states are guided by these initiatives in identifying gaps in legislation, human resource development, and detection infrastructure (IAEA, 2024). The Agency ensures compliance with international obligations under instruments such as the Convention on the Physical Protection of Nuclear Material (CPPNM) and UNSCR 1540 through these frameworks.

Through IAEA assistance in Africa and the Middle East, national radiation detection networks have been established, incident reporting systems have been improved, and national response teams have been established that can coordinate with INTERPOL and customs agencies during trafficking incidents (UNODC, 2023).

6.3 The World Customs Organization and UNODC Role

A cross-border data exchange programme of the WCO strengthens the enforcement of strategic trade controls through a linkage of customs authorities. By using this framework, member states can conduct risk-based inspections on shipments deemed sensitive or high-value (WCO, 2022).

UNODC complements this with its Container Control Programme (CCP), which trains customs officers to identify and interdict illicit goods concealed within legitimate trade routes. By integrating these programs, border agencies can operate under harmonized procedures while respecting national sovereignty while fostering a culture of shared accountability.

6.4 Cooperation between the European Union, ASEAN, and the Middle East

The Schengen Information System II (SIS II) and Frontex Agency for Border and Coast Guard have pioneered integrated border systems in the EU. Through these initiatives, 27 member states can exchange biometric and criminal data almost instantly (Garcia, 2023) showing the benefits of supranational coordination.

ASEAN's Regional Forum on Border Management promotes coordinated maritime surveillance and joint patrols, especially against trafficking routes in the South China Sea. A partnership between the United Arab Emirates and the IAEA and INTERPOL has led to the development of one of the region's most advanced radiation detection infrastructures, including portal monitors, mobile spectrometry units, and command integration at the Jebel Ali Port.

7. Challenges and Ethical Considerations

7.1 Balancing Security, Privacy, and Sovereignty

As border control moves to a data-driven model, there are complex ethical and legal questions to be addressed. The widespread use of biometrics and AI-based surveillance intensifies concerns about data misuse, sovereignty, and privacy. According to Chertoff and Baker (2023), border systems must evolve under transparent governance structures that protect civil liberties without compromising operational effectiveness. Using automated profiling systems too often can lead to discriminatory results, especially if the algorithms are trained with biased datasets.

As a result of international cooperation, states often must share sensitive information with supranational organizations, raising questions about their ownership and jurisdiction. Members of the Schengen Information System (SIS II) of the European Union, for example,

must align their privacy regulations with EU standards, effectively limiting unilateral control over data management (Garcia, 2023) It remains a challenge for global border governance to balance national independence with collective security.

7.2 Digital Divide and Resource Inequality

In addition, there is an unequal distribution of technological capability among nations as another key limitation. In developing and post-conflict countries, there is often a lack of infrastructure, expertise, and financial resources to implement an integrated border management system. The gaps in funding and outdated surveillance equipment leave critical routes vulnerable to exploitation by smuggling networks, particularly in Africa, South Asia, and the Middle East.

Organizations and donors such as the IAEA, UNODC, and the U.S. Department of Energy's Office of Nuclear Smuggling Detection and Deterrence (NSDD) have attempted to bridge this gap through training and technology transfer programs. The sustainability of the program, however, remains an issue, as many nations have difficulties with maintenance costs, staff retention, and procurement delays (IAEA, 2024). It is likely that detection networks will degrade over time if consistent investments are not made, which will undermine regional stability.

7.3 Algorithmic Bias and AI Transparency

In addition to enhancing operational efficiency, AI and machine learning may introduce algorithmic bias and opacity into the decision-making process. Using AI models trained on incomplete or skewed datasets can lead to unjustified detentions or missed threats (Global Security Innovations, 2022). For border operations to remain effective and legitimate, algorithmic transparency and auditability are essential.

The OECD Artificial Intelligence Principles (2021) emphasize the importance of accountability and human oversight in AI applications related to public safety and border control. It is imperative to incorporate ethical design principles into automation such as fairness, explainability, and proportionality to ensure that security goals are met without eroding human rights.

7.4 Cybersecurity and Data Integrity Risks

As border systems become increasingly digital, cybersecurity becomes a defining vulnerability. It is possible for cyberattacks to compromise sensitive data, disrupt logistics, or disable detection systems (UNODC, 2023) if they target customs databases or radiation detection systems. An attack on a port's customs management system, for instance, could delay inspections and obscure illicit shipments.

Countermeasures such as blockchain and quantum encryption are promising, but they require specialized infrastructure and trained personnel to implement. The importance of strengthening cyber hygiene, including staff awareness, access controls, and redundant backups, cannot be overstated.

7.5 Ethical Oversight and International Governance

Technology alone cannot manage global borders; ethical governance frameworks must be incorporated. Multilateral institutions must ensure that emerging tools are used proportionately, lawfully, and transparently. The European Data Protection Supervisor and UN

Human Rights Council audit surveillance systems. Additionally, they provide guidelines for using data responsibly.

The importance of public accountability cannot be overstated. The public can be made more aware of border technologies' deployment and performance through regular reporting. It has been noted by Chertoff and Baker (2023) that the trust deficit between governments and citizens widens as surveillance is expanded without clear oversight or redress mechanisms.

8. Comparative Case Studies

8.1 United States: Layered Technological Integration

Particularly along its southern border, the United States employs a layered border security strategy that integrates physical infrastructure, advanced surveillance, and AI analytics. Unmanned aerial systems (UAS) drones equipped with radars, infrared sensors, and real-time data links have become increasingly useful in monitoring vast and remote terrain, according to Miller and Johnson (2023). As a result of AI-assisted technology, smuggling routes are predicted, and patrol resources are allocated based on satellite imagery, radar data, and ground patrol reports.

To detect nuclear materials, U.S. Customs and Border Protection (CBP) also uses radiation portal monitors and handheld spectrometers. These systems feed data into centralized risk databases accessible by DHS and international partners, exemplifying an integrated data-sharing architecture.

8.2 European Union: Interoperability and Smart Borders

One of the most advanced models of multinational border interoperability is found in the European Union (EU). As the backbone of the EU's digital border system, the Schengen Information System (SIS II), the Entry/Exit System (EES) and the European Travel Information and Authorization System (ETIAS) all play a role. The 27 member states can exchange biometric, visa, and criminal data near-instantly through these systems (Garcia, 2023).

A good example of the EU's focus on automation and predictive security is the Smart Borders Project (European Commission, 2024). A biometric e-gate, advanced passenger screening, and integrated watchlists allow efficiency and control. Considering the EU's General Data Protection Regulation (GDPR), ongoing governance reform is required in order to address privacy concerns and algorithmic fairness.

8.3 Australia: Maritime Border Surveillance and Regional Cooperation

In Australia, maritime surveillance is emphasized, as well as regional collaboration. The Australian Border Force (ABF) uses naval patrols, aerial drones, and satellite tracking as part of its Maritime Border Command (Taylor & Wong, 2021). This framework employs real-time global monitoring to prevent unauthorized boat arrivals and human trafficking.

In addition, Australia participates in the Proliferation Security Initiative (PSI), which targets illicit nuclear and radiological materials transport by sea. As part of its cooperation with Southeast Asian nations, Australia shares maritime intelligence and supports capacity-building initiatives to combat piracy and smuggling.

8.4 United Arab Emirates: Advanced CBRN Detection Infrastructure

As part of a collaboration with the IAEA and INTERPOL, the United Arab Emirates developed one of Middle East's most sophisticated border detection systems. A comprehensive

CBRN detection network is installed at the Jebel Ali Port, one of the world's largest maritime hubs (INTERPOL, 2024) as well as mobile spectrometry units and integrated command-and-control systems.

As a result of strategic investments in technology and international partnerships, the UAE can enhance national security without compromising operational efficiency. In addition to serving as a benchmark for other Gulf states pursuing similar modernization projects, its CBRN network is regionally significant.

8.5 Comparative insights

Throughout these case studies, context matters, since geography, politics, and economics influence each state's approach. As the US and EU focus on technological sophistication and integration, Australia and the UAE are exemplifying regional cooperation and infrastructure-centric strategies. Several programs funded by donors and incremental improvements are prioritized by developing nations.

These examples highlight the importance of flexible, scalable border security architectures that can adapt to shifting global threats while adhering to legal and ethical requirements.

9. Recommendations

- **Adopt an Integrated Detection Architecture**

Under unified command structures, states should develop multilayered systems linking their physical, maritime, and cyber borders. It is essential to incorporate CBRN detection, data fusion centers, and mobile response teams with real-time coordination capabilities into these architectures.

- **Enhance cyber resilience**

Cybersecurity protocols, blockchain for document authentication, and quantum encryption must be incorporated into border security infrastructure. Under the guidance of the IAEA and WCO, international standards for digital integrity should be established.

- **Enhancing international data-sharing mechanisms**

As intelligence-sharing platforms evolve toward automated, AI-enabled systems, anomalies need to be flagged while ensuring privacy is maintained. By harmonizing protocols across INTERPOL, UNODC, and IAEA databases, high-risk movements could be intercepted faster.

- **Integrate ethical and legal oversight**

Ideally, surveillance systems should be operated under transparent data governance frameworks that include independent audits, algorithmic accountability, and grievance mechanisms for individuals whose identities or profiling have been falsely confirmed.

- **Enhance regional cooperation and capacity building**

The donor nations and international agencies should prioritize long-term technical support, training, and equipment maintenance. Sharing of knowledge and harmonizing operations can be fostered through regional cooperation, such as through the African Union or ASEAN.

- **Innovate and invest in research.**

It is imperative that we continue to invest in AI-driven spectral analysis, drone-based detection, and quantum-security border communication systems.

10. Conclusion

The three pillars of border security are intelligence, technology, and trust. The border must become an intelligent ecosystem capable of detecting, deterring, and responding to threats without compromising human rights or economic vitality. In an uncertain geopolitical environment, integrated border management will remain the frontline of defense against illicit trafficking and terrorism.

Achieving global stability requires a shared vision - one that transcends national boundaries, embraces innovation, and upholds ethical principles. It is only through cooperation, transparency, and sustained commitment that the international community can ensure that borders serve as gateways of security and lawful exchange, rather than fault lines of insecurity.

References

1. Anderson, P. (2021) The Politics of Cross-Border Cooperation: Security in a Fragmented World. Global Policy Press. https://www.researchgate.net/publication/358219787_Cross-border_cooperation_a_global_overview
2. Andreas, P. and Snyder, T. (2021) The Politics of Border Security. Oxford University Press. https://doi.org/10.1093/acrefore/9780190264079.013.657
3. Chertoff, M. and Baker, S. (2023) 'Surveillance and Sovereignty: Privacy in the Age of Biometric Borders', National Security Journal, 18(2), pp. 45–67.
4. European Commission (2024) Smart Borders Project: Leveraging Data for Secure Mobility. European Union Publications. https://home-affairs.ec.europa.eu/policies/schengen/smart-borders_en
5. Mamad Chaki, N., & Ashour, M. W. (2021,). Automated border control systems: a literature review. In IET Conference Proceedings CP791 (Vol. 2021, No. 11, pp. 152-157). Stevenage, UK: The Institution of Engineering and Technology.
6. Gonzalez, A. and Lee, R. (2024) Pre-emptive Security through Data Sharing: Advanced Passenger Information Systems and Global Risk Mitigation. International Border Policy Review, 14(2), pp. 100–118. https://www.europeanpapers.eu/e-journal/schengen-european-borders-introduction-special-section#:~:text=Abstract:%20The%20functioning%20of%20the,of%20the%20Schengen%20Borders%20Code.
7. IAEA (2024) Illicit Trafficking Database Annual Report. Vienna: International Atomic Energy Agency. https://www.iaea.org/sites/default/files/25/03/itdb-factsheet.pdf
8. ICMPD (2024) Integrated Border Management Strategy Review. International Centre for Migration Policy Development, Brussels. https://www.icmpd.org/
9. INTERPOL (2024) Global Border Management Programme Annual Report. Lyon: INTERPOL Publications. https://www.interpol.int/content/download/22267/file/INTERPOL%20Annual%20Report%202023%20EN.pdf
10. International Security Technology Review (2023) Drones and Automated Surveillance in Border Management: Global Applications and Trends. International Institute for Security Studies.
11. Lee, R. and Nakamura, H. (2024) 'Blockchain in Border Security: Transforming Customs Processes', International Journal of Digital Commerce, 9(2), pp. 88–104. https://doi.org/10.1007/978-3-031-52419-6_6
12. Miller, T. and Johnson, R. (2023) 'Technological Advancements in U.S. Border Security: Integrating Technology and Human Oversight', Security Studies Review, 28(4), pp. 45–62.
13. Patel, S. and Nguyen, H. (2024) 'Aerial Surveillance and the Future of Border Security: The Role of Drones in Integrated Systems', Journal of Security Technology and Policy, 15(2), pp. 88–104. https://doi.org/10.24235/itej.v9i2.126

14. Taylor, P. and Wong, S. (2021) 'Maritime Border Control in Australia: Strategies and Controversies', International Maritime Security Quarterly, 19(3), pp. 155–172. https://www.ussc.edu.au/maritime-security-strategies-for-japan-and-australia-report-of-the-inaugural-jade-fellows
15. UNODC (2023) Global Report on Trafficking of Nuclear and Radiological Materials. United Nations Office on Drugs and Crime, Vienna. Https://www.unodc.org/unodc/data-and-analysis/world-drug-report-2023.html
16. WCO (2024) Cross-Border Data Exchange for Strategic Trade Controls. Brussels: World Customs Organization. https://mag.wcoomd.org/magazine/strategic-trade-control-enforcement-project-a-new-component-of-the-wco-compliance-and-enforcement\package/#:~:text=Key%20identified%20challenges&text=Until%20the%20high%2Dlevel%20issues,liability%20and%20seized%20property%20disposition.

Education and Training in Nuclear Security in Sudan – Status and Prospects

Bokhari Ahmed Moneir Ahmed

University of National and World Economy, Sofia, Bulgaria
bokhari.ahmed@unwe.bg

Abstract: This paper examines the current status and future prospects of education and training in nuclear security in Sudan. It emphasizes the critical role of human resource development in establishing a sustainable nuclear security regime, ensuring that professionals are qualified to prevent, detect, and respond to threats involving nuclear and radioactive materials. The study reviews Sudan's historical evolution of higher education, national efforts toward nuclear infrastructure development, and existing academic and training initiatives in nuclear security. It identifies major challenges, including limited infrastructure, lack of specialized instructors, budget constraints, and brain drain. Drawing on international best practices, particularly those promoted by the IAEA, the paper provides recommendations for enhancing nuclear security education, research, and technical training. Strengthening national capacity through targeted education, modernized training facilities, and international collaboration is essential to achieve a robust and sustainable nuclear security framework in Sudan.

Key words: Nuclear security, Education, Training, Human resource development, Sudan, IAEA, Higher education.

1. Introduction

Education and training play an essential role in ensuring that experts and competent authority in nuclear security are prepared and qualified to work in nuclear security tasks this include analyses national nuclear security needs, to prevent, detect and combat the threat of sabotage and the use of nuclear and radioactive material for malicious acts, and to prepare effective response measures to nuclear security events. Human resource development is the process of improving individual, group, and organizational performance through training, education, shared experience, and organizational development initiatives. (Nadler & Nadler, 2012). It emphasizes the need for comprehensive training programs that incorporate a systems approach to nuclear security to keep nuclear material secure and protected. Assesses the impact of professional development courses on nuclear security educators, emphasizing the need for sustainable communities of practice

However, the nuclear security regime builds on the implementation of relevant international legal instruments information protection cybersecurity physical protection, material accounting and control and national response plans, and contingency.so there are Multi-specialty in nuclear security should be consider, so the measures are vital is to prevent, detect and respond to nuclear security events or other malicious, theft and sabotage of nuclear and radioactive materials to protect persons, property, society and the environment from the harmful consequences of nuclear security events (IAEA -2011).

Sudan has made significant efforts to build a national nuclear infrastructure, it has maintained international nuclear security standards, and is prepared to commence nuclear energy development. While Sudan's engagement with nuclear technology has mostly focused

on peaceful applications, such as medical and agricultural uses, this study aims to explore the existing educational and research landscape related to nuclear security in the country. Sudan has been an IAEA Member State since May 31, 1957, Sudan passed its first nuclear law in February 2017; law puts Sudan as one of only five African countries with a complete nuclear law that integrates radiation safety, nuclear security, and safeguards. Furthermore, the law established a separate regulatory agency by legally establishing the Sudanese Nuclear and Radiological Regulatory Authority (SNRRA).

According to an International Atomic Energy Agency (IAEA) team, Sudan has strengthened its national nuclear security regime by establishing substantial nuclear security systems and security measures. Sudan has prepared detailed studies on the human resources required for a nuclear power programme and should implement the activities identified to support the development of qualified and competent staff for the key organizations. (INIR-sudan-2018).

Sudan, while not possessing a nuclear programme, has shown interest in peaceful applications of nuclear technology. This paper investigates the extent to which education and research in nuclear security have been pursued in Sudan, and outlines the prospects for future advancements in this area. In summary, this paper provides an overview of the current status of education and training in nuclear security in Sudan and outlines prospects for further development.

2. Historical Evolution of Higher Education in Sudan

Higher education in Sudan has evolved through distinct phases, each marked by a pivotal moment that has profoundly influenced its trajectory, with degrees ranging from Bachelor's to Doctorate level, well structured into three tiers comprising Bachelor's and Master's degrees.

Inception and Early Growth (1902)

The roots of formal English education in Sudan can be traced back to 1902 with the establishment of the Gordon Memorial College (GMC). Originally conceived as a primary school, GMC laid the foundation for formal English education in Sudan. Evolving into a college over time, it played a crucial role in shaping the early landscape of higher education in the country.

Emergence of Khartoum University (1951)

The year 1951 witnessed the merger of Khartoum University College and the University of London, resulting in the creation of a new institution. Graduates from Khartoum College were conferred with Bachelor's degrees from the University of London. By 1956, this institution underwent a transformation, becoming the esteemed University of Khartoum and assuming a central role in Sudanese higher education (Hassan, 2020).

Institutional Regulation (1975)

A significant stride in the regulation of higher education occurred in 1975 with the enactment of a dedicated law. The establishment of the National Council for Higher Education marked a pivotal moment, ushering in a unified institution tasked with overseeing higher education policies and coordinating the activities of its constituent institutions (Hibatallah & Shukran, 2020).

Transformational Shift in 1990

The Sudanese higher education landscape experienced a transformative shift in 1990. The advent of the National Salvation "revolution" in June 1989 brought about a corresponding

"revolution" in higher education during the same year. This period witnessed substantial changes and reforms in the higher education system, aligning with the broader political transformations occurring in the country (Hibatallah & Shukran, 2020).

The historical narrative reveals a dynamic evolution of higher education in Sudan, spanning from the establishment of the first formal English education institution to the formation of Khartoum University, the regulatory measures in 1975, and the transformative changes in 1990. These milestones collectively delineate the growth and development of higher education in Sudan over the years. Nuclear Security education and Training in Sudan

In the country of Sudan, there are numerous higher education institutions, both public and private, that offer undergraduate nuclear technology programs. The academic landscape encompasses Over 40 universities offering a wide range of disciplines and graduate programs, including engineering, science, computer science, physics, chemistry. In spite of the fact that a number of universities offer courses on nuclear physics and engineering, there is a noticeable absence of a distinctive focus on nuclear security in the educational curriculum. The University of Sudan University of Sciences and Technology (SUST), with its Department of Nuclear Engineering, is a notable exception since it initiated planning in 2006 and welcomed its inaugural class in 2009. As a result of the pioneering program, offering a BSc degree in nuclear engineering, 110 qualified nuclear engineers will be able to contribute to the workforce of the Sudan Nuclear Power Plant (NPP) and other nuclear activities in the country.

There is currently no course available at the public universities that deals with this important aspect of nuclear security despite its overarching importance. A distinguished institution in Sudan, the Sudan Academy of Sciences (SAS) was established in 2004 under the Ministry of Science and Technology and is currently under the Ministry of Higher Education and Scientific Research is a well-known fact that SAS offers three.

Master's degree programs, namely Medical Physics, Radiation Protection, and Nuclear Science, locally. SAS offers three Master's degrees locally, namely Medical Physics, Radiation Protection, and Nuclear Science

Through training initiatives, the Sudan Atomic Energy Commission (SAEC) fosters expertise in addition to academic endeavors. Various training programs, seminars, and workshops are conducted by the SAEC within the Radiation Safety Institute (RSI). In addition to academics and research, on-the-job training emphasizes practical skills.

The Training Centre is a well-established hub for nuclear-related training courses, thanks to significant support from the International Atomic Energy Agency (IAEA). Collectively, these institutions are contributing to the development of Sudan's nuclear workforce and the advancement of nuclear knowledge.

Table 1: shows the universities in Sudan with nuclear programs

University	Programe	Started date
University of Sudan University of Sciences and Technology (SUST)	BSc in nuclear engineering	2009
Sudan Academy of Sciences (SAS)		2004
Karari University	Porpoise Master in Nuclear Security	

3. Institutional and Regulatory Framework

The Sudanese Nuclear and Radiological Regulatory Authority (SNRRA) serves as the national regulatory body mandated by the *Nuclear and Radiological Regulatory Act of 2017*. It oversees all activities involving nuclear and radioactive materials, including authorization, inspection, enforcement, and emergency preparedness. According to the presentation delivered at the *First Regional Workshop on Harmonizing Regulatory Requirements for RPO Qualification and Competence* (Ali Sid Ahmed- 2024), Sudan has developed a structured framework for defining the qualifications and responsibilities of Radiation Protection Officers (RPOs) and Qualified Experts (QEs). The national system follows the IAEA Safety Standards GSR Part 3 and IAEA Education and Training Guidelines, ensuring that RPOs are designated within each authorized facility to implement radiation protection programs and ensure compliance with safety and security requirements. QEs are responsible for technical oversight in areas such as calibration, dosimetry, and radiation measurement.

The SNRRA's framework establishes detailed competence requirements, including academic qualifications, practical experience, and periodic retraining. These qualifications are verified through a structured licensing process that aligns with IAEA recommendations. However, the 2023 conflict has interrupted this framework, resulting in halted inspections, suspension of training programs, and limited regulatory communication with operators. Re-establishing the RPO and QE training and certification system will be critical for restoring Sudan's national nuclear security infrastructure and ensuring regulatory resilience. (Ali Sid Ahmed-2024).

4. Nuclear Security Support Centers of Excellence in Sudan

Sudan has actively sought technical assistance from the International Atomic Energy Agency (IAEA) for the peaceful applications of nuclear technology in various development and service sectors. This assistance has been channeled through the IAEA's technical cooperation and nuclear security department.

In collaboration with the IAEA, Sudan has established Nuclear Security Support Centers to provide comprehensive training, technical advice, and support to nuclear security experts. One of these centers is located at the Sudanese Nuclear and Radiological Regulatory Authority, under the supervision of the Secretary General of SNRR.

Since 2003, Sudan has been actively participating in IAEA regional nuclear security initiatives. Staff from the SNRRA and other relevant organizations have engaged in 23 regional security training events, impacting approximately 80 individuals through IAEA regional nuclear security programs (2003–2015). Additionally, 13 national training courses aligned with IAEA Nuclear Security Series events have been meticulously planned. Sudanese participants further enriched their knowledge by attending a regional workshop on nuclear security culture in Ghana (June 2018), an international nuclear security training session for emerging nations in Russia (May 2018), and a course in Jordan (June 2018).

These initial efforts have laid a solid foundation for the expansion of nuclear security training and technical support across various national organizations. The beneficiaries include front-line officers, first responders, emergency response personnel, intelligence and law enforcement authorities, trainers, nuclear regulators, and policymakers, covering diverse aspects of nuclear security.

The sustained effectiveness of nuclear security hinges on ongoing national efforts, including technology development, human resource development, education, and training. The nucleus of these efforts resides in a center whose core activities are inherently technical and

scientific. This center focuses on providing training related to the use, calibration, and maintenance of equipment, coupled with raising awareness about nuclear security issues within all competent authorities. Since its establishment, the center has successfully hosted several training sessions for various government agencies.

The courses offered by the center encompass vital aspects of nuclear security, including physical protection of nuclear material and facilities, safeguards, accountancy, and control, the use of radiation detection equipment by government officials, safety and security associated with radioactive sources, radiation detection equipment and maintenance, combating illicit trafficking of nuclear and radiological materials, search and recovery of orphan sources, and awareness-raising for policymakers and other public officials requiring a nuanced understanding of nuclear security.

5. Challenges of Education and Training in Nuclear Security

Many challenges are facing our country in education and training in nuclear security. The training to be effective and efficient, more advanced technology labs will be required, which will be equipped with the necessary tools. Simulators and animations should be used to make the training process more effective and efficient. Additionally, new challenges in nuclear security require the introduction of new courses that will help manage them effectively.

The availability of qualified teachers to cover most, if not all, areas of nuclear security because of nuclear security many disciplines need different specialists with highly competent knowledge. In order to provide some expensive instruments and equipment, and with few types of equipment with low-trained enforcement agents, the country needs to collaborate with developing countries. As a result of moving and migrating scholars and experts within Sudan, it has been estimated that 18 scientists with a PhD degree, 4 professors, and over 50 master's degree holders have migrated within the last few years, thus putting their expertise and knowledge at risk.

Lastly, the fourth challenge that needs to be addressed is the lack of well-equipped laboratories that could be used in the educational setting and are readily available in the nuclear science or engineering departments. The best laboratories for this purpose would be those already available in these departments. It is true, however, that few of these organizations are equipped with the specialized equipment and tools that may be needed for a comprehensive educational programme on nuclear security.

Furthermore, international cooperation and collaboration are essential for a nuclear security program to be successful. For this reason, it is necessary to establish contacts with international universities that offer programs in nuclear security through student/research exchange programs. Moreover, in order to diversify the educational curriculum of more public universities, nuclear security should be introduced into their existing programs among the growing number of public universities (Noreen Iftakhar- 2017).

6. Impact of the Ongoing Conflict on Nuclear Security Education and Human Resource Development

The outbreak of armed conflict in Sudan in April 2023 has severely disrupted national institutions, including those responsible for education, research, and nuclear security. The escalation of violence in Khartoum and other regions has caused extensive damage to physical infrastructure, displacement of professionals, and interruption of university operations. Institutions such as the Sudan University of Science and Technology (SUST), the Sudan Atomic Energy Commission (SAEC), and the Sudan Academy of Sciences (SAS) have all

faced closures or significant limitations in their activities. Laboratories and training facilities that supported nuclear education and research are no longer fully operational. Many students and faculty members have been displaced, and communication between academic departments and national authorities has been severely constrained. As a result, the pipeline for developing new experts in nuclear security has been critically affected. In addition to halting academic progress, the conflict poses indirect risks to nuclear and radiological safety.

The disruption of regulatory oversight and physical protection measures increases vulnerability to theft, loss, or misuse of radioactive sources used in medical and industrial applications. Ensuring their security under unstable conditions requires urgent coordination between the Sudanese Nuclear and Radiological Regulatory Authority (SNRRA), the IAEA, and relevant international partners. Rebuilding the education and training ecosystem will require targeted post-conflict recovery programs focusing on restoring damaged infrastructure, providing scholarships and remote learning opportunities for displaced students, and re-establishing institutional links with the IAEA and other international training networks. A national strategy for human resource development in nuclear security should also incorporate contingency planning to ensure continuity of education and regulatory functions during crises.

7. International Education and Training in Nuclear Security

Several efforts have been made in the past to improve education and training because of multilateral cooperation. The International Atomic Energy Agency (IAEA) has an essential role. It offers a variety of services and materials, such as expert missions, training programmes, and guide publications. The IAEA developed technical guidelines for a Master of Science (M.S.) and certificate programme in nuclear security to help member nations to build and maintain their nuclear security capabilities.

Through the International Nuclear Security Education Network (INSEN) that was established in 2010, the IAEA, academic and research institutions, as well as other stakeholders, work together to promote long-term nuclear security education, (INSEN, 2010). It was only two years later that the IAEA established the International Network for Nuclear Security Training and Support Centres (NSSC Network) in February 2012 to facilitate sharing of information and resources and to promote coordination and collaboration among countries.

Moreover, the IAEA strives to offer a comprehensive training programme that is interactive and participatory in order to sustain human resource development in nuclear security and it developed its first eLearning module in the area of radiation detection. The objective of this computer-based learning tool is to provide front-line officers with a good basic understanding of why, when and how radiation detection is performed at borders, in addition to many eLearning modules in nuclear security transport security, computer security, nuclear material accounting and control for security purposes, radiological crime scene management

In 2023, IAEA opened the Nuclear Security Training and Demonstration Centre (NSTDC) at the IAEA's laboratories in Seibersdorf, Austria, which will further support countries in their efforts to enhance their nuclear security expertise. (IAEA, 2023).

WINS Academy offers a professional development programme and hosts a variety of workshops, training courses, roundtables and webinars annually in locations around the world to improve and expand of understanding of nuclear security. Enrollment in the WINS Academy was encouraged for participants through the scholarship fund of the WINS Academy. Many students have registered at WINS academy for nuclear security-related courses as a result of the occurrences. Certification programme, and a small number of Sudanese are the alumni of WINS Academy.

In Africa, nuclear security education is in its early stages of institutionalization. Regional capacity-building efforts are supported by the African Regional Cooperative Agreement for Research, Development, and Training Related to Nuclear Science and Technology (AFRA) and its educational branch, AFRA-NEST, which conducts workshops and faculty development initiatives. Countries such as Ghana, Kenya, and Sudan have benefited from IAEA Technical Cooperation projects, which provide training in nuclear safety, radiation protection, and physical security (IAEA, 2020).

At the University of National and World Economy (UNWE) in Bulgaria, nuclear security education is delivered through a multidisciplinary framework that integrates policy, law, and management aspects. In 2015, UNWE initiated a Master's in Nuclear Security program with support from the IAEA to help students meet emerging threats and regulatory challenges while adhering to international nuclear security standards (UNWE, 2023). The university collaborates with European and Bulgarian international institutions to promote best practices. It serves as a model for institutions in developing regions seeking to enhance nuclear security education. Globally, progress in nuclear security education has strengthened national and regional frameworks. However, disparities in access, infrastructure, and sustainability remain.

8. Competency Requirements for Nuclear Security

The first step is to create a comprehensive list of all requirements. The process of creating a competency framework typically involves a series of steps. As shown in the Fig. 2.

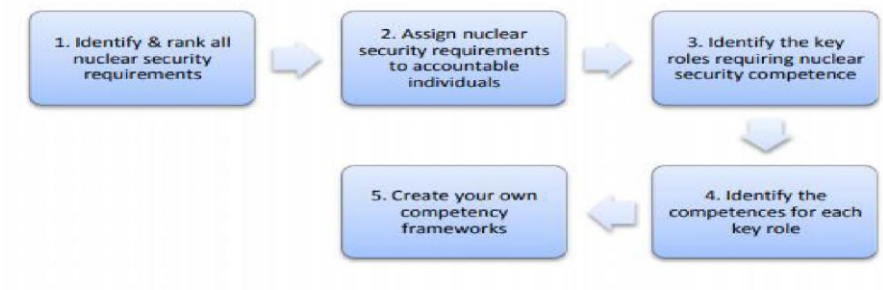


Fig. 2. Competency framework Source: (IAEA - 2013)

Analysis of Competence Gaps

When developing competence, it is important to make an assessment to the workforce when devolving nuclear security personnel, the competence gap analysis enables the regulatory body and operator to compare existing competences of its staff to the required competences to the job description. IAEA has published a number of safety standards and other publications in which the need and importance of ensuring regulatory competence are emphasized and provides guidance for training of the regulatory staff, including the training needs. Accordingly, the report IAEA" Systematic Assessment of Regulatory Competence Needs (SARCoN)" provides a methodology for two essential complementary elements of the competence management process, developing competence profiles and conducting a competency gap analysis. In this regard, all competent authorities should make Competence gap analysis for all nuclear security officers and identify the need to carry out their responsibility (IAEA-2015a).

Competency Model

Several approaches have been used by the nuclear industry in the development of training programmes based on job/task/competency analyses. The significance of IAEA efforts for devolving training the competency model presented in this section groups competencies into discrete categories, depicting them in terms of four quadrants or Systematic Assessment of Regulatory Competence Needs (SaRCON). The model will be used to identify the present nuclear security competence and future need for the regulatory body or competent authority and operator. This approach will help to identify gaps in competencies that would be filled by either training programmes or recruitment.

Table 2: The four-parts competency model presented applied to a regulatory

Competences related to the legal ● Regulatory and Organizational Basis ● Legal Basis ● Regulatory Policies and Approaches ● Regulations And Regulatory Guides ● Management System	**Technical disciplines competences** ● Basic science and technology ● Applied science and technology ● Specialized science and technology in security
Competences related to a regulatory ● Review and assessment ● Authorization ● Inspection ● Enforcement ● Development of regulations and guides. ● Security plan and response	**Personal and behavioral competences** ● Analytical thinking and problem solving ● Personal effectiveness and self-management ● Communication ● Teamwork ● Managerial and leadership competences ● Security culture

Source: IAEA 2013 https://www-pub.iaea.org/MTCD/Publications/PDF/Pub1635_web.pdf

9. Conclusion

The state of nuclear security education and training in Sudan reveals both progress and vulnerability. While the country has benefited from sustained cooperation with the International Atomic Energy Agency (IAEA), the African Regional Cooperative Agreement (AFRA), years of instability and the recent conflict have severely disrupted academic and regulatory progress. Universities, laboratories, and training centers have been damaged or rendered inoperative, and many qualified experts have been displaced, creating a critical shortage of skilled professionals in the field.

Despite these challenges, Sudan stands at a pivotal moment of opportunity. Post-war recovery offers the chance to rebuild a stronger, more integrated framework for nuclear security education, one that aligns with international standards while addressing national realities. Rehabilitating damaged educational and research infrastructure, restoring the national qualification framework for nuclear security.

Future progress will depend on rebuilding institutional capacity and promoting sustainable human resource development through partnerships with international and regional centers of excellence. The integration of digital learning tools, targeted scholarships, and exchange programs can accelerate recovery and reduce the long-term impact of brain drain.

Sudan's commitment to nuclear security education and training should be a cornerstone of its national recovery and development strategy. The Sudanese government can re-build a competent and resilient nuclear security workforce capable of ensuring the safe, secure, and peaceful application of nuclear technology, both within and across its borders, through strategic investment, international collaboration, and policy coherence.

Recommendations

During the recent conflict, Sudan faced severe disruptions to the nuclear security education and training sector. As a result of the war, educational infrastructure has been destroyed, academics have been displaced, and research and training programs have been interrupted. To build a resilient, sustainable nuclear security education and training system, immediate recovery measures must be combined with long-term strategic reforms. To strengthen the national framework for human resource development and nuclear security capacity building, the following recommendations are proposed:

Curriculum Enhancement and Academic Reform

Developing comprehensive, multidisciplinary curricula that include nuclear security, non-proliferation, radiation protection, and nuclear forensics should be a priority for Sudanese universities. By combining technical and policy-oriented components, a workforce capable of addressing evolving nuclear security challenges. In addition to fostering regional knowledge exchange, collaboration with the International Nuclear Security Education Network (INSEN) is essential to guarantee compatibility with international standards.

Post-War Restoration of Educational Infrastructure

Rehabilitation of damaged laboratories, training facilities, and university campuses is a top priority for the city of Khartoum capital city. Institutions such as the Sudan University of Science and Technology (SUST) and the Sudan Academy of Sciences (SAS) should be supported to restore their capacity to deliver specialized programs in nuclear engineering, radiation protection, and nuclear security. Partnerships with the IAEA and regional organizations can provide technical and financial assistance for rebuilding facilities and re-equipping laboratories.

Reintegration of Displaced Academics and Students

Conflict has forced many qualified professionals and students to relocate. The government, in collaboration with international partners, should establish reintegration programs, research fellowships, and visiting scholars' opportunities to re-engage Sudanese experts. Virtual learning platforms and temporary international placements can ensure education continuity during the transition period.

Investment in infrastructure and technology

Modern laboratories, simulation equipment, and e-learning systems are critical to revitalizing nuclear security education. Investment should prioritize digital transformation, including the development of national online training platforms and the use of IAEA and WINS eLearning modules. This approach will increase accessibility, particularly while physical reconstruction continues (WINS, 2022).

Strengthening the Policy and Legal Framework

In order to promote nuclear security education, Sudan is required to establish its legal and policy foundations. To develop standards for academic programs, licensing, and professional certifications, the Sudanese Nuclear and Radiological Regulatory Authority (SNRRA) should coordinate with universities and the Ministry of Higher Education. Sustainability and accountability will be enhanced by integrating nuclear security into the broader post-conflict reconstruction plan within a cohesive national education recovery framework.

Increasing funding and international support

The Sudanese government should prioritize human resources development, with support from the IAEA, AFRA, and donor states. Rebuilding educational capacity and expanding training require financial resources. Funds should be allocated for faculty training, scholarships, and infrastructure repairs. Sudan's rebuilding process can be accelerated by targeting international funding mechanisms, as was the case with Iraq and Ukraine after the conflict.

Collaboration, Knowledge Sharing, and Regional Partnerships

Collaboration with leading institutions is crucial for expertise and innovation such as the University of National and World Economy (UNWE), Developing partnerships with leading institutions can facilitate research collaborations, faculty exchanges, and capacity building activities. A revitalization of regional cooperation through AFRA-NEST and the African Union's scientific networks is also necessary to ensure Africa's progress is shared.

Continuous Professional Development and a Security Culture

Developing a robust nuclear security culture requires lifelong learning and continuous professional development. Sudan should institutionalize refresher and certification programs through cooperation with the WINS Academy and IAEA eLearning platforms. Incentives for ongoing professional education will ensure that personnel across the regulatory, academic, and operational sectors remain competent and up to date with global nuclear security practices (WINS, 2022).

Promoting Brain Gain and Talent Retention

Rebuilding after a conflict requires reverse academic migration. Sudan should launch a "brain gain" initiative to attract and retain experienced scientists and educators through competitive research grants, tenure opportunities, and international collaborations. Reinvesting in local expertise will be the key to sustainable nuclear education and research capacity.

References

1. Ali Sid Ahmed - (SNRRA). (2024). Sudanese Nuclear and Radiological Regulatory Authority National Framework for RPO and QE Qualification and Competence: Presentation at the First Regional Workshop on Harmonizing Regulatory Requirements for RPO Qualification and Competence. Cairo: SNRRA.
2. Hibatallah Mohammad Elhassan Salim, Shukran Abd Rahman (2020 Change and development of higher education in Sudan - IIUM Repository (IRep). http://irep.iium.edu.my/81308/
3. IAEA (2011). Nuclear Security Recommendations on Nuclear and Other Radioactive Material Out of Regulatory Control.
4. IAEA .2022 IAEA Mission to Sudan Finds Improved Nuclear Security Infrastructure, Encourages Strengthening of Nuclear Security Detection and Response ,accessed BY 01.01.2023 https://www.iaea.org/newscenter/pressreleases/iaea-mission-to-sudan-finds-improved-nuclear-security-infrastructure-encourages-strengthening-of-nuclear-security-detection-and-response.

5. IAEA 2013 safety reports series no. 79 managing regulatory body competence https://www-pub.iaea.org/MTCD/Publications/PDF/Pub1635_web.pdf
6. Iftakhar, N., & Noor, S. (2017). Nuclear Security Education and Training in Pakistan. International Journal of Nuclear Security, 3(1). https://doi.org/10.7290/ijns030108
7. International Nuclear Security Education Network (INSEN). (n.d.). IAEA. https://www.iaea.org/services/networks/insen
8. Mission Report on the Integrated Nuclear Infrastructure Review (INIR) Phase One Mission. (2018). https://www.iaea.org/sites/default/files/documents/review-missions/inir-sudan-030918.pdf
9. Mohammad, H. & Abdul Rahman, S. (2020). Change and Development of Higher Education in Sudan. PESA Uluslararası Sosyal Araştırmalar Dergisi, 6 (2), 113-120. Retrieved from https://dergipark.org.tr/en/pub/pesausad/issue/56282/647747
10. Nadler, Z., & Nadler, L. (2012, August 21). Designing Training Programs. Routledge.
11. The IAEA Nuclear Security Training and Demonstration Centre is Set to Open. (n.d.). https://www.iaea.org/newscenter/news/the-iaea-nuclear-security-training-and-demonstration-centre-is-set-to-open

Application of Machine Learning in the Financial Sector

Kameliya Petkova and Melani Lyubenova

University of National and World Economy, Sofia, Bulgaria
kpetkova_23183002 @unwe.bg, mlyubenova_23183005@unwe.bg

Abstract. This paper examines the main applications of machine learning in the financial sector, focusing on customer service, fraud detection, credit risk assessment, marketing, pricing, and income estimation. It discusses how data-driven models can support more accurate predictions, improve operational efficiency, and enable more personalized financial services. Examples from academic research illustrate the potential of machine learning in financial product recommendations, financial planning, and mortgage risk assessment. The paper also considers limitations related to data quality, algorithmic bias, privacy, and transparency. It concludes that responsible use of machine learning can improve financial decision-making, provided that institutions combine technological innovation with appropriate human oversight and regulatory safeguards.

Keywords. Machine learning, Financial services, Credit risk, Fraud detection, Financial inclusion

1. Introduction

In today's rapidly changing technological landscape, machine learning (ML) has become an increasingly important and widely discussed topic. As a branch of artificial intelligence (AI), machine learning enables computers to learn from data and improve their performance without being explicitly programmed for every task. Its applications are transforming many industries by helping companies improve their processes, automate tasks, and make data-driven decisions. Machine learning has therefore become an important tool for businesses seeking to remain competitive and innovative. This paper examines how the financial sector benefits from machine learning and how it can improve financial services and operations. Machine learning is no longer a niche technology used only by large technology companies. Today, businesses of different sizes and across various industries use it to explore new opportunities and increase efficiency. With advances in data processing and the rapid growth of available data, organizations increasingly recognize its value for informed decision-making, personalized customer experiences, and more efficient operations. Financial institutions must decide where and how machine learning can have the greatest impact on their activities. Whether it is used to streamline internal processes, improve financial forecasts, or strengthen risk assessment systems, its potential applications are extensive.

The aim of the paper is to examine the main applications of machine learning in the financial sector, focusing on customer service, fraud detection, credit risk assessment, marketing, pricing, and income estimation.

2. A Brief History of Machine Learning

The history of machine learning dates back to the 1950s, when Alan Turing introduced the idea of machines capable of thinking and Arthur Samuel developed a program that learned to play checkers, one of the earliest examples of a machine learning system. Between the 1960s

and the 1980s, researchers explored fundamental algorithms such as perceptrons and nearest-neighbor methods, although progress was limited by the computing power available at the time. In the 1990s, more effective algorithms, including support vector machines and decision trees, emerged and allowed machine learning to outperform traditional statistical methods on certain tasks. The twenty-first century brought rapid growth, driven by big data and improved hardware, while ensemble methods such as random forests became increasingly popular. From the 2010s onward, deep learning transformed the field, enabling breakthroughs in image and speech recognition and establishing machine learning as a central part of modern AI.

3. Differences Between Machine Learning and Artificial Intelligence

Because of its ability to learn and make decisions, machine learning is often described as artificial intelligence (AI), although it is actually a subset of AI. Until the late 1970s, it developed as part of the broader evolution of artificial intelligence before becoming a more distinct area of research. Machine learning and artificial intelligence are closely related but different fields within computer science.

Machine learning is a part of AI that allows machines to learn from data and improve over time without being explicitly told what to do in every situation. In other words, all machine learning falls within AI, but not all AI involves machine learning. AI may include applications such as robots and voice assistants, while machine learning focuses mainly on identifying patterns in data to make predictions or decisions.

AI is the broader field concerned with developing systems that imitate aspects of human intelligence, including reasoning, decision-making, and problem-solving. ML, by contrast, focuses on training systems to recognize patterns in data and improve their performance over time without detailed, task-specific programming.

4. Applications of Machine Learning in the Financial Sector

Banking and finance are among the sectors adopting machine learning to analyze transaction records, customer interactions, and internal documents. These methods can help institutions identify patterns, understand customer needs, and support operational decisions. Recent reviews describe applications in financial risk management and credit scoring [1].

With machine learning, financial institutions can improve fraud detection, strengthen credit scoring, support regulatory compliance, and increase customer engagement. This paper presents some of the main areas in which machine learning is used and explains how it is changing the sector.

4.1. Improving Customer Service

Machine learning (ML) is increasingly used in finance to predict individual preferences, anticipate customer needs, and offer personalized products and services. By analyzing large volumes of information, including transaction histories, spending patterns, and demographic data, financial institutions can provide a more tailored experience that improves customer satisfaction and loyalty.

For example, by examining transaction histories and demographic information, ML models can predict which financial products may interest a particular customer. The study [2] proposed an adaptive collaborative filtering algorithm with personalized time-decay functions for recommending financial products. By taking changing customer preferences into account, this approach improves recommendation accuracy compared with traditional methods.

ML models can also anticipate future customer needs by analyzing behavioral data and identifying early signs of changing preferences. This proactive approach allows financial institutions to offer suitable products and services even before customers express a specific interest.

Vanguard's reinforcement learning model for financial goal planning uses deep reinforcement learning to identify appropriate savings levels across multiple goals and income sources [3]. By anticipating future needs and adjusting strategies accordingly, this approach helps customers work toward their financial goals.

Machine learning also enables financial institutions to tailor their offerings to individual customers, so that products and services better reflect their needs and preferences. Bank of America introduced Erica, an AI-powered virtual financial assistant that provides personalized financial guidance. Erica uses natural language processing and predictive analytics to offer proactive suggestions based on individual customer information.

4.2. Fraud Detection and Prevention

Financial fraud involves deceptive practices intended to obtain money or other financial benefits unlawfully. In finance, this includes credit card fraud, identity theft, account takeovers, and insurance fraud. The growing volume of digital transactions increases exposure to such activities and creates a need for more advanced detection methods.

Traditional fraud detection often relies on rule-based systems that apply predefined criteria to identify suspicious activity. Although these systems can recognize known fraud patterns, they often struggle with new or changing tactics.

Machine learning uses algorithms trained on historical data to identify patterns that may be difficult to capture with fixed rules. A systematic review of financial fraud detection studies documents the use of classification and other ML techniques, while emphasizing the importance of datasets and evaluation methods [4].

Financial institutions are increasingly introducing real-time fraud detection systems powered by ML algorithms. These systems analyze transactions as they occur to identify and prevent suspicious activity promptly. Deep learning models such as long short-term memory (LSTM) networks can capture relationships over time and help detect more complex fraudulent behavior.

4.3. Automation and Operational Efficiency

- *Risk Management*

Put simply, a Machine learning model is a mathematical prediction model whose performance can improve as it learns from more data. Applying machine learning requires two main elements: a model and data. There are many different types of ML models. A model describes the process used to generate predictions from the available data.

The following discussion considers four main types of machine learning models relevant to credit risk modeling. In addition to choosing a model, a substantial amount of historical data is needed. These data must include both the outcome to be predicted, known as the target, and other variables related to that outcome, known as features. In credit risk modeling, historical information is collected for many individuals. The target
indicates whether a person has defaulted, while features may include income, age, and employment status.

Once the data have been collected, the model is trained using historical examples. Machine learning models are designed to identify relationships between the features and the

target. After training, the model can make predictions for new cases where the features are known but the outcome is not.

Maintaining financial stability requires responsible lending and sufficient regulatory capital to absorb losses arising from defaults. Classification methods such as logistic regression (LR) have long been standard tools for estimating default probability, but they face growing operational demands and regulatory requirements. Advances in machine learning now allow models to process larger datasets and potentially achieve higher predictive accuracy. Algorithms such as decision trees (DT), support vector machines (SVM), and neural networks (NN) can improve credit risk modeling, although they are not always straightforward to implement. A recent review discusses the benefits and limitations of machine learning for consumer credit risk modeling [5].

Lending is one of the central activities of the banking system. Financial institutions generate income by providing loans to businesses and individuals, who must repay the borrowed amount with interest. However, lending exposes the institution to the risk that a borrower will fail to meet their obligations. This is known as credit risk and is one of the main risks faced by banks.

Credit risk deserves particular attention in financial risk management because it is among the most significant threats facing the sector.

Credit risk modeling is a form of economic forecasting that aims to identify which potential borrowers are likely to default. In this setting, machine learning can reduce prediction errors and support better decisions, including lending decisions.

A customer's credit risk refers to the possibility that the borrower will fail to meet financial obligations, such as repaying a loan. A central component is default risk: the probability that the borrower will fail to repay part or all of the loan. Accurate assessment and management of this risk are essential because they directly affect a financial institution's profitability, regulatory compliance, and long-term stability.

Traditionally, credit risk assessment has relied on rule-based systems and statistical models using a limited set of variables, such as income, credit history, and outstanding debt. These approaches may struggle to capture complex, nonlinear relationships between different factors. This is where machine learning can make an important contribution.

Machine learning models can analyze large amounts of structured and unstructured data to identify patterns related to creditworthiness. Some of the main models used are:

- Logistic regression - a traditional but still widely used method that estimates the probability of default. It often serves as a baseline in credit risk modeling because its results are relatively easy to interpret.
- Decision trees and random forests – models that work well with tabular data and can capture nonlinear relationships and interactions between features. Random forests can offer strong accuracy and stability.
- Gradient boosting algorithms (such as XGBoost and LightGBM) - ensemble methods that build predictive models in stages. They are widely used because of their strong performance on tasks such as default prediction.
- Neural networks – particularly deep learning models, which are increasingly used with high-dimensional or unstructured information, such as transaction histories or text from loan applications.
- Support vector machines (SVM) and k-nearest neighbors (k-NN) – less commonly used in practical credit risk modeling, but potentially effective in particular situations, especially with smaller datasets.

These models are trained on historical data that may include repayment behavior, credit utilization, demographic characteristics, and external factors such as economic indicators. Once trained, they can estimate the probability of default for new or existing customers.

- *Credit Risk in Real Estate*

This area includes risks associated with mortgage loans and commercial real estate lending. In the study [6], a deep learning model was developed using data on more than 120 million US mortgages issued between 1995 and 2014. The model included loan-specific and macroeconomic variables, as well as information at ZIP-code level. Its purpose was to predict the likelihood of prepayment, foreclosure, and different states of delinquency. The main findings included:

- Nonlinear relationships: The model revealed strong nonlinear links between the variables and borrower behavior, particularly in decisions to repay mortgages early.
- Sensitivity to macroeconomic factors: State-level unemployment rates had considerable explanatory power, suggesting a close relationship between mortgage markets and the wider economy.
- Heterogeneous effects: Economic conditions affected different borrower groups in different ways, highlighting the need for more tailored risk assessments.

This approach provides more accurate and detailed predictions than traditional linear models.

Climate-related risks can also affect mortgage portfolios. In principle, combining property-level exposure information with borrower and loan characteristics could help institutions assess these risks; however, the accuracy of such models depends on the availability and quality of the underlying data.

Commercial real estate investments present additional risks, including changes in occupancy, market conditions, and financing costs. Machine learning may help analyze these factors, although the usefulness of any model needs to be demonstrated with appropriate data and independent evaluation.

In the study [7], a model based on spatio-temporal modeling is proposed. This model combines tree boosting with a latent spatio-temporal Gaussian process to account for nonlinear relationships and variation across locations and time. Their findings suggest improvements in predicting mortgage default probabilities and portfolio losses.

4.4. Marketing

- *Lead Scoring*

Machine learning models analyze customer data, including demographics, transaction histories, and online behavior, to estimate the likelihood that a person will respond to a marketing campaign. This can help financial institutions focus their resources on customers who are more likely to be interested in a particular offer. Predictive models should nevertheless be evaluated on representative data before being used in practice.

- *Customer Segmentation and Profiling*

Machine learning algorithms group customers according to shared characteristics, allowing financial institutions to tailor marketing strategies to different segments. RFM analysis (recency, frequency, and monetary value) can be combined with predictive models to support customer profiling and sales planning. The effectiveness of these methods depends on the quality and relevance of the available data.

- *Customer Churn Prediction and Retention Strategies*

Machine learning models analyze customer behavior to identify signs that a customer may leave, allowing financial institutions to take action before that happens. By examining behavioral data, ML models can predict which customers are likely to end their relationship with an institution and suggest tailored retention measures, such as personalized offers or better customer service. These measures aim to strengthen loyalty and reduce customer loss.

4.5. Price Elasticity

- *Interest Rate Forecasting*

Accurately forecasting interest rate movements is important for understanding their potential effect on demand for financial products. Models such as long short-term memory (LSTM) networks and autoregressive integrated moving average (ARIMA) models can be applied to historical interest rate data. Their performance depends on the forecasting horizon, data quality, and changing market conditions.

- *Pricing Optimization*

Machine learning techniques can help financial institutions examine customer behavior, market conditions, and other relevant information when setting prices for financial products. However, pricing decisions also depend on funding costs, regulation, competition, and institutional risk policies. Model outputs therefore need to be assessed alongside conventional financial analysis.

- *Pricing Interest Rate Derivatives*

Interest rate derivatives, such as swaps and options, are important tools for managing exposure to interest rate fluctuations. Machine learning can be explored as a complementary approach to derivative valuation by modeling complex relationships in financial data. Such applications require careful validation against established pricing methods and market observations.

4.6. Communication Optimization / Next Best Action

Machine learning has become an important tool in finance, particularly for improving communication strategies and identifying the *next best action* (NBA) for customer engagement. By analyzing large amounts of data, ML models can predict customer behavior, personalize interactions, and support better decisions.

Financial institutions can use machine learning to examine customer preferences and choose suitable communication channels, such as telephone, email, or digital messaging. This approach may make campaigns more relevant, although any claimed improvement in campaign performance needs to be supported by documented results.

AI-based next-best-action recommendation systems help financial institutions develop personalized marketing and sales strategies. By analyzing customer preferences and behavior, AI can suggest effective ways to engage each customer, potentially improving loyalty and increasing revenue.

Machine learning can also support personalized investment services by analyzing financial goals, risk tolerance, and portfolio characteristics. Automated investment platforms may use such information to suggest portfolio allocations or adjustments. Recommendations should remain consistent with the client's circumstances and applicable investor-protection requirements.

Integrating machine learning into communication optimization and next-best-action strategies is changing the financial sector. By enabling more personalized interactions, supporting decision-making, and increasing customer engagement, ML can improve operational efficiency and strengthen customer relationships.

4.7. Income Estimation

In finance, accurately estimating an individual's income is essential for credit scoring, loan approval, and financial planning. Traditional methods mainly rely on documented sources of income, such as payslips and tax returns. However, many people, particularly those working in the gig economy or informal employment, do not always have access to such documents.

Machine learning offers ways to estimate nontraditional or irregular income using alternative data, including transaction histories. Such approaches may support assessments where conventional income documentation is limited, but they require careful attention to privacy, accuracy, and fairness.

Another approach involves analyzing human mobility patterns. A study published in Applied Geography examined the use of movement data to estimate income distribution. By identifying patterns in spatial activity and travel routes, ML models can infer socioeconomic status and income levels, particularly in areas where conventional data are unavailable.

Super-apps, which combine services such as messaging, payments, and shopping, generate large amounts of user data. Research suggests that analyzing behavior within these applications can improve income estimation models. Spending habits, transaction frequency, and patterns of service use provide additional context that may not be captured by traditional financial information.

Although ML provides useful tools for income estimation, potential bias in these models requires attention. Bias can arise from sample selection, measurement errors, or algorithm design. For example, some algorithms may place greater weight on features with more clearly defined values, potentially disadvantaging people with less conventional sources of income.

Accurate income estimation using ML can also support financial inclusion. By drawing on alternative data, financial institutions can assess the creditworthiness of people without a traditional credit history, including immigrants and young adults. This can help people who have previously had limited access to financial services participate more fully in the economy.

Machine learning is changing income assessment in the financial sector through alternative data and advanced analytical methods. These developments can improve estimates of nontraditional income while supporting greater fairness and financial inclusion. As ML models continue to develop, ethical use and efforts to reduce bias will be essential to achieving fair outcomes.

5. Conclusion

The integration of machine learning (ML) into finance is changing how institutions operate, interact with customers, and manage risk. From predictive customer service to income estimation and regulatory compliance, ML offers new opportunities for personalization, efficiency, and more detailed analysis.

In fraud detection and prevention, ML models can identify unusual activity and adapt to changing patterns in real time, helping reduce financial crime. Automation has simplified routine tasks such as loan approval and customer service, lowering costs and improving response times. Risk management is also becoming more proactive as ML systems can forecast market movements and assess creditworthiness using both traditional and alternative data.

In direct marketing and communication optimization, ML helps tailor messages and communication channels, improving customer engagement and retention. Advanced recommendation systems identify suitable products or next steps, supporting revenue growth while offering customers more relevant services. Similarly, the ability to analyze spending behavior and demand trends improves price elasticity analysis and helps financial products remain competitive and responsive to market changes.

However, the use of ML also presents challenges. Estimating income in less clearly documented situations, such as gig work and informal employment, raises questions about data quality and fairness. Legal and ethical issues also deserve serious attention. Algorithmic bias, transparency, personal data protection, and accountability are central to building trust and complying with regulations such as the GDPR and the EU AI Act.

Overall, machine learning has become an important part of modern finance, encouraging innovation, improving services, and supporting better decisions. To realize its full potential, however, the financial sector must adopt responsible AI practices that balance technological capabilities with ethical standards and legal accountability.

Acknowledgments

This work has been carried out in the framework of the Project "Application of Artificial Intelligence for Sustainable Business Development in the Context of Climate Change", financially supported by the Science Fund of the University of National and World Economy, Bulgaria (Grant No. NID NI 35/2025/A).

References

1. Tian, X., Tian, Z., Khatib, S. F. A., Wang, Y. (2024). Machine learning in internet financial risk management: A systematic literature review. PLOS ONE, 19(4), e0300195. https://doi.org/10.1371/journal.pone.0300195
2. Ghiye, A., Barreau, B., Carlier, L., Vazirgiannis, M. (2023). Adaptive collaborative filtering with personalized time decay functions for financial product recommendation. In Proceedings of the 17th ACM Conference on Recommender Systems, pp. 798–804. Association for Computing Machinery. https://doi.org/10.1145/3604915.3608832
3. Mohammed, S., Bealer, R., & Cohen, J. (2021). Embracing advanced AI/ML to help investors achieve success: Vanguard reinforcement learning for financial goal planning. arXiv. https://arxiv.org/abs/2110.12003
4. Hernandez Aros, L., Bustamante Molano, L. X., Gutierrez-Portela, F., Moreno Hernandez, J. J., Rodríguez Barrero, M. S. (2024). Financial fraud detection through the application of machine learning techniques: A literature review. Humanities and Social Sciences Communications, 11, Article 1130. https://doi.org/10.1057/s41599-024-03606-0
5. Montevechi, A. A., Miranda, R. de C., Medeiros, A. L., Montevechi, J. A. B. (2024). Advancing credit risk modelling with machine learning: A comprehensive review of the state-of-the-art. Engineering Applications of Artificial Intelligence, 137, Article 109082. https://doi.org/10.1016/j.engappai.2024.109082
6. Sirignano, J., Sadhwani, A., Giesecke, K. (2016). Deep learning for mortgage risk. arXiv. https://arxiv.org/abs/1607.02470
7. Kündig, P., & Sigrist, F. (2024). A spatio-temporal machine learning model for mortgage credit risk: Default probabilities and loan portfolios. arXiv. https://arxiv.org/abs/2410.02846