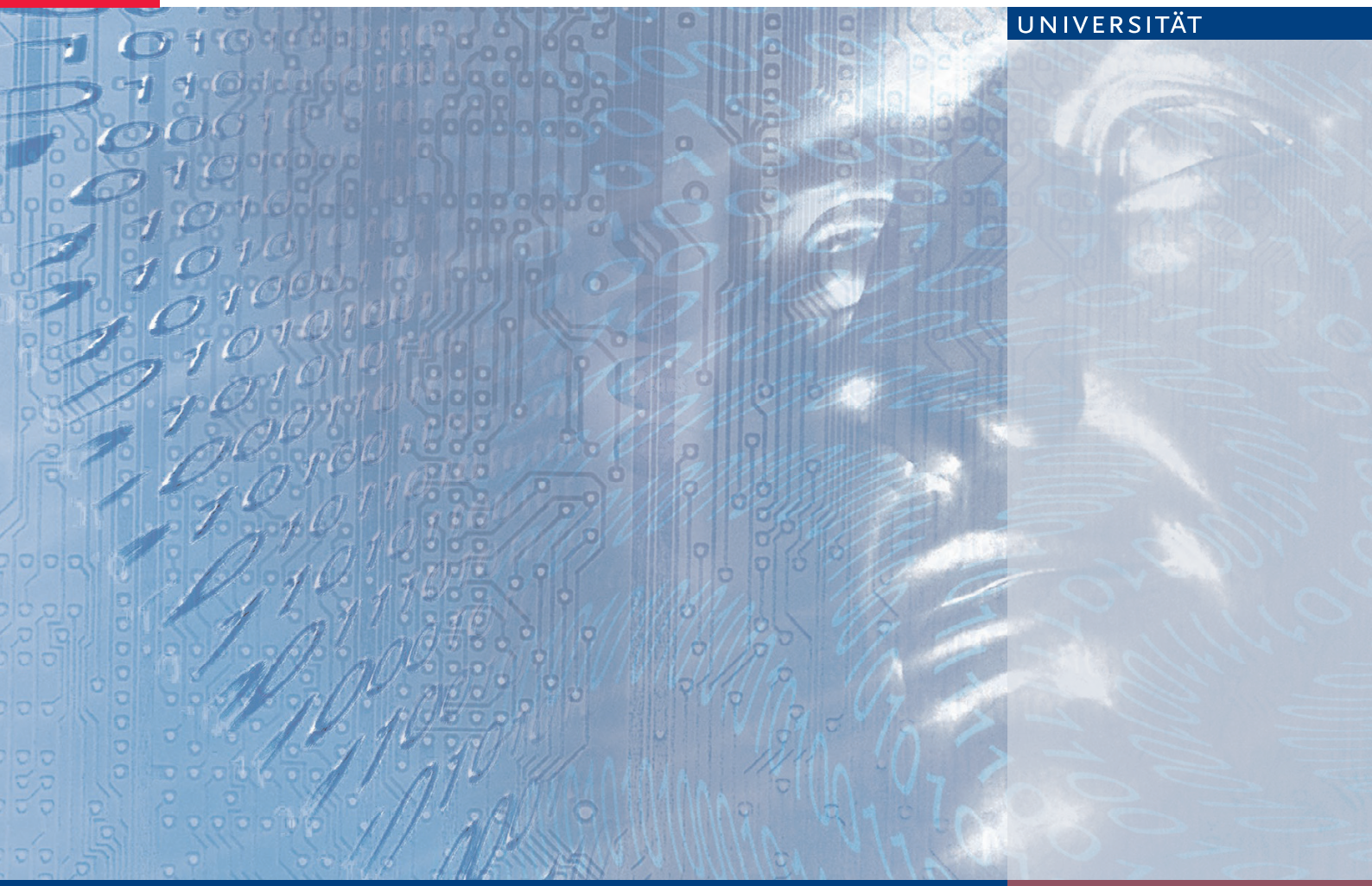
TRAUNER VERLAG
UNIVERSITÄT



SCHRIFTENREIHE INFORMATIK
55

DOUCEK PETR ■ SONNTAG MICHAEL ■
TKÁČ MICHAL ■ JANOŠKOVÁ MÁRIA (EDITORS)

IDIMT-2026

Resilience: ICT, Business and Society

34th Interdisciplinary Information Management Talks
Sept. 2–4, 2026
Hradec Králové, Czech Republic

JKU
JOHANNES KEPLER UNIVERSITÄT LINZ

TRAUNER VERLAG
UNIVERSITÄT

Schriftenreihe
Informatik

55

Doucek Petr ■ Sonntag Michael ■
Tkáč Michal ■ Janošková Mária (Editors)

IDIMT-2026

Resilience: ICT, Business and Society

34th Interdisciplinary
Information Management Talks
Sept. 2–4, 2026
Hradec Králové, Czech Republic

Impressum

Schriftenreihe Informatik

Doucek Petr ■ Sonntag Michael ■
Tkáč Michal ■ Janošková Mária (Editors)
IDIMT-2026
Resilience: ICT, Business and Society
34th Interdisciplinary Information Management Talks

This publication was partially supported by the Bratislava University of Economics and Business, Faculty of Business Economics with seat in Košice (Slovakia), the Johannes Kepler University Linz (Austria), and the Prague University of Economics and Business (Czech Republic) through institutional support for the long-term conceptual development of science and research at the Faculty of Informatics and Statistics (IP400040). The doctoral section of the conference was additionally supported through project IG409035.

The Conference IDIMT-2026
took place September 2–4, 2026
in Hradec Králové, Czech Republic

Programme Committee
Delina Radoslav, SK
Doucek Petr, CZ
Janošková Mária Ria, SK
Lisnik Anton, SK
Maryška Miloš, CZ
Nedomova Lea, CZ
Neubauer Georg, AT
Pavliček Antonín, CZ
Pitner Tomáš, CZ
Pucihar Andreja, SI
Rainer Karin, AT
Schoitsch Erwin, AT
Shaaban Abdelkader, AT
Sonntag Michael, AT
Tkáč Michal, SK

© 2026
The Author(s) 2026

Herausgeber:
Bratislava University of Economics and Business
Dolnozemská cesta 1
852 35 Bratislava
Slovak Republic

Kommisionsverlag:
TRAUNER Verlag + Buchservice GmbH
Köglstraße 14, 4020 Linz
Österreich/Austria

Herstellung:
Dataform Media GmbH
Julius-Raab-Straße 8,
2203 Großebersdorf
Österreich/Austria

PRINTED IN AUSTRIA

DOI: 10.35011/IDIMT-2026

ISBN 978-3-99195-218-3
www.trauner.at

TABLE OF CONTENS

NEW PATHS TO TACKLE COMPLEX CRISIS: DECISION SUPPORT BY AI AND DECENTRALIZED DATA MANAGEMENT

EVIDENCE-BASED DECISION SUPPORT IN HEALTH CRISES: AI-ENHANCED KNOWLEDGE MANAGEMENT, GOVERNANCE, AND AI LITERACY 15
Melissa Hagendorn, Selina Bahardoust, Karin Rainer, Jennifer Rao, Alois Leidwein, Drazen Ignjatovic, Georg Neubauer, Anna Sumereder

CHANGE MANAGEMENT MODELS IN THE AGE OF INDUSTRY 4.0: ARE TRADITIONAL APPROACHES STILL RELEVANT AND IN USE? 25
Lukáš Klarner, Petr Řehoř, Jaroslav Vrchota, Petra Matoušková, Kamila Němečková, Monika Maříková, Ladislav Rolínek

DECISION SUPPORT FOR THE MONITORING AND MANAGEMENT OF BATHING WATER QUALITY 33
Georg Neubauer, Talha Anwar, Martin Jung, Verena Parzer, Bernd Reichl

CYBERRESILIENCE FOR SECURITY

CRYPTOGRAPHY IN A POST-QUANTUM WORLD 43
Michael Sonntag

MULTIPLE ETHICAL ASPECTS AND LIMITS OF ICT RESILIENCE

SELECTED ETHICAL ASPECTS AND LIMITS OF ICT RESILIENCE 55
Anton Lisnik

ARCHITECTURE OF RESPONSIBILITY FOR GENERATIVE AI IN SOCIAL MEDIA: BEHAVIOURAL DATA, PROFILING AND ETHICAL-ALGORITHMIC RISK 63
Krzysztof Krusiec, Andrzej Palej

WHEN AI IS DISCLOSED: A PILOT EXPERIMENTAL STUDY OF PERCEIVED AUTHENTICITY, TRUST AND PERSUASIVE INTENT OF SHORT TEXTS 73
Tomáš Sigmund

SOCIAL MEDIA BEHAVIOUR AS TRAINING DATA FOR ARTIFICIAL INTELLIGENCE: ETHICAL CHALLENGES AND RISKS 81
Ivan Katrenčik, Monika Zatrochová

DATA, AI AND DIGITAL-DRIVEN TRANSFORMATION: SHAPING SUSTAINABLE DIGITAL FUTURES

DYNAMIC DIGITAL CAPABILITIES FOR DIGITAL RESILIENCE OF SMES: CONCEPTUAL FOUNDATIONS AND RESEARCH DESIGN 91
Andreja Pucihar, Mirjana Kljajic Borstnar

DIGITAL MATURITY OF SMES AS A CATALYST FOR SUSTAINABLE DIGITAL FUTURES: A MULTI-CRITERIA ANALYSIS OF EU MEMBER STATES 99
Eva Ardielli

MARKET DYNAMICS AND RISKS IN THE GLOBAL MACHINE LEARNING INDUSTRY .. 107
Maria Bartekova

STRATEGIC ASCENDANCE OF INFORMATION TECHNOLOGY IN BUSINESS STRATEGY AND PERFORMANCE 115
Martin Zbořil

RESILIENCE FOR AUTONOMOUS SYSTEMS

RE-THINKING AUTOMATION: INTENSIVE HUMAN-MACHINE COLLABORATION 127
Erwin Schoitsch

HEURISTIC-GUIDED DETECTION OF RESOURCE EXHAUSTION VULNERABILITIES FOR ENHANCING RESILIENCE IN AUTONOMOUS SYSTEMS 135
Loui Al Sardy, Mamdouh Muhammad, Abdelkader Magdy Shaaban, Mahmoud Baklizi

AN INTEGRATED AI ASSURANCE AND RISK MANAGEMENT FRAMEWORK FOR AI-ENABLED TOOLS AND DIGITAL WORKSPACES IN AIMS5.0 143
Abdelkader Magdy Shaaban, Christoph Schmittner

DATA GOVERNANCE, DATA ANALYTICS & AI

FROM "GARBAGE IN → MYSTERY OUT" TO TRUSTWORTHY ANALYTICS AND AI 155
Miloš Maryška, Petr Doucek, Felix Espinoza

RAG-BASED MARKET INTELLIGENCE ASSISTANT FOR BUSINESSES 167
A Sujith Kumar, Shalini Aggarwal, Prerna Dhawan, Erik Boltun, Roland Weiss, Kiran Sood, Simon Grima

RELATIONSHIP BETWEEN DATA GOVERNANCE AND ARTIFICIAL INTELLIGENCE GOVERNANCE 175
Martin Zelenka

SOCIAL MEDIA, CREATIVITY AND INNOVATION: SYNERGIES AND TENSIONS

FROM RANKING TO REMOVAL: AI-DRIVEN SELECTION, MODERATION AND GOVERNANCE IN SOCIAL MEDIA 185
Antonín Pavlíček, František Sudzina, Jana Syrovátková, Richard A. Novak

USER BEHAVIORAL FACTORS INFLUENCING INFORMATION SECURITY ON SOCIAL NETWORK SITES 197
Michael Anderle, Martin Lukáš

BETWEEN CREATIVITY AND MANIPULATION: HOW CZECH ONLINE NEWS FRAMES AI-GENERATED CONTENT, PLATFORMS, AND RESPONSIBILITY 205
Lukas Herout, Anna Ulrichova

FROM PARTICIPATORY CREATIVITY TO CO-INNOVATION: THE ROLE OF SOCIAL MEDIA AND AI IN INNOVATION ECOSYSTEMS 213
Lucie Rožmberská

RE-EXAMINING THE LINK BETWEEN CHRONOTYPE AND SOCIAL NETWORKING SITES USE FREQUENCY 221
František Sudzina, Antonín Pavlíček

ESTIMATION OF THE TRUTH OF NEWS AND THEIR VERIFICATION (SURVEY) 229
Jana Syrovátková, Antonín Pavlíček

BETWEEN TECHNOLOGICAL SIMPLIFICATION AND DIGITAL COLLABORATION: EVIDENCE ON SOLO AUTHORSHIP IN COMPUTER SCIENCE FROM THE DBLP DATABASE IN THE YEARS 2000–2024 237
Lubomír Štěpánek

AI & DATA-DRIVEN SUPPLY CHAIN MANAGEMENT

GOVERNANCE OF SOCIAL VALUE AI FOR PUBLIC PROCUREMENT AND SUPPLY CHAINS 247
Radoslav Delina, Michaela Sulova, Patrik Koba

OPERATIONAL EFFICIENCY OPTIMIZATION OF SUPPLY CHAIN 259
Sharfraj Ahmed, Shalini Aggarwal, Prerna Dhawan, Kiran Sood, Simon Grima, Erik Weiss

GENERATIVE AI IN ESG COMMUNICATION AND SUSTAINABLE CORPORATE BRANDING: TRUST AND SYNTHETIC AUTHENTICITY 267
Martin Bachura, Magdaléna Freňáková, Tatiana Jurovčáková, Slávka Bocánová

IDENTIFYING INNOVATION IN PUBLIC PROCUREMENT: A TEXT MINING ANALYSIS OF V4 COUNTRIES 275
Miroslava Barkóciová, Mariana Ivaničková

DETERMINANTS OF TAX RELIABILITY AMONG SLOVAK CONSTRUCTION SMES 283
Marian Brontvaj, Robert Verner

THE TRUST PHENOMENON IN AI-DRIVEN MARKETING: ESG AS A MITIGATING TOOL 291
Mária Dolná, Magdaléna Freňáková, Jozefína Hvastová

HETEROGENEITY OF REGISTERED SOCIAL ENTERPRISES IN PUBLIC PROCUREMENT: A DESCRIPTIVE ANALYSIS ON SLOVAK DATA 299
Abraham Gocze, Gabriel Demeter

GENERATIONAL ASYMMETRY IN AI ADD-ON LAYER IMPLEMENTATION FOR ERP: A LONGITUDINAL CASE STUDY IN A SPECIFIC MANUFACTURING ENTERPRISE 307
Vladislav Kašpar

COMPOSITION EFFECTS IN THE UTILITY GREENIUM 315
Daniel Kratky, Robert Verner

LEAN AND RARE: ENVIRONMENTAL AWARD CRITERIA IN CENTRAL EUROPEAN TENDERING 323
Jakub Sieber, Erik Boltun

WHEN ONE OFFER IS ALL THERE IS: PROCUREMENT SAVINGS IN OBSERVED SINGLE-OFFER TENDERS 331
Michal Tkáč, Michal Tkáč, Alžbeta Džujková

FROM CAMPUS TO COMPANY: COLLABORATIVE INNOVATION IN THE AGE OF AI

FROM CAMPUS TO COMPANY – COLLABORATIVE INNOVATION IN THE AGE OF AI . 341
Radan Baše, Ondřej Čuka, Ondřej Skřehota, Miroslav Jaroš, Tomáš Pitner, Oldřich Trenz

BUSINESS, TRADE AND MARKETING IN THE AGE OF AI

AI IN SUSTAINABLE BRANDING: CONSUMER BEHAVIOUR IN THE ORGANIC PRODUCTS MARKET 351
Mária Janošková, Erik Weiss, Jozef Zuzik

THE IMPACT OF EMAIL MARKETING ON THE PERFORMANCE OF E-COMMERCE BUSINESSES IN THE DIGITAL ECONOMY 359
Adela Feranecová, Alena Nováková

CONSUMER BEHAVIOUR OF GEN Z IN THE AGE OF AI 367
Jaromír Veber, Petra Králová

COMPETENCE DEMAND IN AI-DRIVEN LABOUR MARKET IN SLOVAKIA 375
Cecília Olexová, Patrik Koba

LIFELONG LEARNING AS A DRIVER OF SUSTAINABLE COMPETITIVENESS IN SELECTED CEE COUNTRIES 383
Cecília Olexová, Kristína Herzigová, Natália Olexa

EARLY CAREER & STUDENT SHOWCASE

ARTIFICIAL INTELLIGENCE ADOPTION IN SMALL AND MEDIUM-SIZED ENTERPRISES: THE MODERATING EFFECT OF FIRM SIZE IN V4 COUNTRIES 393
Sabina Janikovičová, František Pollák

AIROF: A PRACTICAL FRAMEWORK FOR ADAPTIVE RISK AND OPPORTUNITY MANAGEMENT IN THE USE OF AI TOOLS IN CONSULTING SMES..... 401
Ludmila Jiříčková

ARTIFICIAL INTELLIGENCE AND TECHNO-STRESS AT WORKPLACES: A SCOPING REVIEW 409
Kamila Němečková, Petr Řehoř, Nikola Soukupová, Lukáš Klarner

SELECTED PERSPECTIVES ON REPUTATIONAL RISK IN THE CONTEXT OF CYBERSECURITY 417
Romana Sobotová, František Pollák

PROJECT MANAGEMENT, STRATEGIC MANAGEMENT AND THEIR INTERCONNECTION IN PRACTICE..... 425
Hana Theer Vítková, Hana Mohelská

ANNEX

STATEMENT OF THE PUBLICATION ETHICS AND PUBLICATION MALPRACTICE 435
LIST OF AUTHORS..... 437

SPONSORS OF IDIMT 2026

CONTINUING THE CONVERSATION

The 34th edition of the Interdisciplinary Information Management Talks (IDIMT) marks a new organisational chapter in the history of the conference. The Bratislava University of Economics and Business, Faculty of Business Economics with seat in Košice, has joined the consortium of organising institutions and has assumed responsibility for the coordination of the conference and its proceedings, while the long-standing academic cooperation with the Prague University of Economics and Business and Johannes Kepler University Linz continues. Although the organisational responsibilities evolve, the principles that have shaped IDIMT over more than three decades remain unchanged.

Since its establishment in 1993, IDIMT has provided an international forum where researchers from different disciplines meet to exchange ideas, challenge assumptions and discuss the implications of emerging developments in information management and related fields. Throughout its history, the conference has continuously adapted to new scientific questions and technological developments. Annual themes have naturally evolved to reflect the challenges of their time, while the conference has preserved its distinctive interdisciplinary character and its commitment to open academic dialogue.

As the conference enters this new organisational stage, our intention is not to redefine what IDIMT represents, but to continue building upon the values that have earned the confidence of its community over many years. We remain committed to scientific quality, rigorous peer review, international cooperation and interdisciplinary discussion. We also believe that the continued strength of IDIMT lies in its ability to bring together researchers from different disciplines, institutions and countries to examine contemporary challenges from multiple perspectives. New themes will continue to emerge, yet the conference will remain guided by the same principles of openness, scholarly curiosity and mutual respect that have characterised it since its beginnings.

This continuity would not have been possible without the dedication of the many individuals and institutions that have contributed to the conference throughout its history. We express our sincere appreciation to its founders, former organisers, members of the Programme Committee, reviewers, authors and participants who have collectively shaped IDIMT into a respected international scientific community. Their work has created a strong foundation upon which future generations can continue to build.

Institutions evolve. Technologies transform. Generations succeed one another. Yet one thing remains unchanged. Understanding grows when we talk. Perhaps this is why, even after more than three decades, the Interdisciplinary Information Management Talks remain as relevant as ever.

Michal TKÁČ

Petr DOUCEK

Michael SONNTAG

NEW PATHS TO TACKLE COMPLEX CRISIS: DECISION SUPPORT BY AI AND DECENTRALIZED DATA MANAGEMENT

EVIDENCE-BASED DECISION SUPPORT IN HEALTH CRISES: AI-ENHANCED KNOWLEDGE MANAGEMENT, GOVERNANCE, AND AI LITERACY

Melissa Hagendorn, Selina Bahardoust, Karin Rainer, Jennifer Rao, Alois Leidwein

Austrian Agency for Health and Food Safety GmbH
melissa.hagendorn@ages.at, selina.bahardoust@ages.at, karin.rainer@ages.at

Drazen Ignjatovic, Georg Neubauer

Austrian Institute of Technology GmbH
drazen.ignjatovic@ait.ac.at, georg.neubauer@ait.ac.at

Anna Sumereder

BOC Products & Services AG
anna.sumereder@boc-group.com

DOI: 10.35011/IDIMT-2026-15

Keywords

Artificial intelligence; decision support; health crises; knowledge management; AI literacy; governance.

Abstract

In health crises, informed policy decisions depend on the timely integration of diverse and rapidly changing evidence. Potentially relevant information may be distributed across scientific, institutional, operational, and societal sources, while decision-makers must often act before these inputs can be fully consolidated into usable evidence structures. Artificial intelligence (AI) may help public institutions manage evidence more dynamically. Yet its value depends on more than technical capability. Drawing on Austrian research on pandemic decision support and AI-enhanced evidence formation, this paper examines how AI-supported knowledge management can strengthen evidence infrastructure for decision support in health crises. It discusses how knowledge graphs, retrieval-augmented generation, bounded AI agents, and decentralized data infrastructures may support traceable evidence use and more transparent alignment between policy measures, affected groups or systems, and intended outcomes. At the same time, AI use in decision support raises legal, ethical, and operational risks, including bias, opacity, accountability gaps, data-quality limitations, and overreliance. Addressing these risks requires transparent data practices, expert validation, human-centred governance, and institutional AI literacy. Rather than merely accelerating decision support, these safeguards allow AI to strengthen evidence infrastructure for decision-making in health crises by supporting traceability, accountability, and trustworthiness.

1. Introduction: Decision-making under crisis conditions

Health crises confront policy-makers with the need for urgent decisions under incomplete and rapidly changing conditions. During epidemics and pandemics, decisions on countermeasures must account not only for epidemiological developments, but also for healthcare capacities, vulnerable groups, legal constraints, social acceptance and compliance, public trust, economic consequences, and unintended side effects. Public health policy is therefore not a purely technical optimisation problem, but a balancing act under uncertainty (WHO, 2021).

The central challenge extends beyond a lack of reliable and relevant data to the transformation of fragmented and context-dependent information into usable evidence for policy decisions. Crisis situations generate information captured in diverse sources, such as scientific literature, surveillance systems, public health agencies, international organisations, legal documents, operational reports, and societal feedback. Human-maintained evidence bases, literature reviews, and decision-support workflows can struggle to keep pace with fast evolutions. Hence, the limitation of static decision support lies not only in the absence of AI, but also in its reliance on manually updated knowledge structures in dynamic crisis environments.

AI-enhanced decision support may help address this limitation by retrieving, condensing, updating, and structuring relevant information from various sources (SAPEA, 2025). It may also support more transparent links between policy objectives, possible measures, expected effects, and unintended side effects. However, AI-supported outputs may also be incomplete, biased, insufficiently contextualised, or affected by limitations in data quality and reliability (Cheong, 2024; Panteli et al., 2025). These risks make AI literacy relevant from the outset, because responsible use depends on whether human actors can interpret AI-supported outputs, recognise uncertainty and bias, and understand where technical support ends and accountable human judgement begins.

AI cannot resolve normative trade-offs, determine political priorities, or remove institutional responsibility. Its role must remain supportive: AI can improve the informational conditions under which human policy-makers decide, but it cannot replace accountable decision-making. This paper examines how AI-supported evidence pathways may contribute to responsible decision support in health crises. It argues that such systems require not only technical performance, but also traceable knowledge management, data infrastructures, human-centred governance, and institutional AI literacy (European Parliament and Council of the European Union, 2024; Long & Magerko, 2020). The aim is therefore not simply to introduce AI into crisis management tools, but to develop evidence pathways that remain transparent, accountable, trustworthy, and usable for human decision-makers.

The following sections illustrate this argument through a project pathway from a web-based pandemic decision-support demonstrator towards AI-enhanced dynamic evidence formation.

2. ROADS to health: Evidence-based decision support in health crises

The ROADS to Health project served as a starting point with the aim to develop a web-based demonstrator for evidence-based decision support in epidemic and pandemic management (ROADS to Health, 2023). The demonstrator allowed users to describe key features of a crisis situation, such as pathogen characteristics, affected or vulnerable groups, and relevant contextual conditions, and to define strategic aims for pandemic management. Based on this input, it linked these aims to possible measures, their expected effects, and potential unintended side effects.

To build this logic, national and international evidence was collected, and pandemic-related parameters were systematically catalogued, including epidemiological, social, cultural, political, and

infrastructural factors. These inputs were used to develop a catalogue of strategic, tactical, and operational objectives, possible measures, and associated consequences. The demonstrator was then developed, tested, and refined in dialogue with stakeholders from crisis management, science, and critical infrastructure to ensure that the concept reflected practical decision-making needs.

ROADS to Health therefore showed how pandemic decision support can be structured around the relationship between crisis characteristics, policy aims, measures, effects, and side effects. However, it remained a demonstrator rather than a fully operational decision-support system. Its development also revealed a central limitation: maintaining and updating evidence structures manually is resource-intensive, particularly in dynamic health crises where evidence, recommendations, and contextual conditions evolve rapidly. This limitation provides the starting point for AI-Health-Path, which investigates how AI can support dynamic evidence formation, automated data consolidation, and more transparent alignment between measures, objectives, expected effects, and unintended side effects.

3. AI-Health-Path: From manual evidence management to AI support

Building on this foundation, AI-Health-Path explores how AI-enhanced knowledge management can reduce the manual burden of maintaining pandemic decision-support systems while preserving traceability, human responsibility, and expert validation (AI-Health-Path, 2025).

The study explores two central functions: (1) the automated consolidation of relevant data and evidence; (2) the AI-supported mapping of measures to objectives in epidemic and pandemic management. Its aim is to improve data quality and coverage, while making the relationships between objectives, measures, expected effects, and unintended side effects more transparent and traceable with targeted processes. Domain-specific stakeholders, including representatives from the Austrian Agency for Health and Food Safety (AGES) and the Austrian Ministry of Health (BMASGPK), contribute scientific, regulatory, and practical perspectives. AI-Health-Path identifies process steps in which automation may be useful, particularly where large volumes of heterogeneous information need to be screened, structured, compared, or updated. The study also distinguishes between types of data according to their suitability for automation. Some sources can be processed more directly because regulatory guidelines or quality standards are available, whereas others require expert assessment or medical quality control before they can responsibly serve as trustworthy decision support. Relevant sources include pandemic plans, recommendations from World Health Organization (WHO) and European Centre for Disease Prevention and Control (ECDC), standard operating procedures, scientific publications, meta-studies, and open data publications (ECDC, 2025; WHO, 2024). AI may support this work through semantic annotation, vectorised search, large language model-generated summaries, and machine-learning-based comparison of historical data.

In this context, the purpose of AI is not to replace expert judgement, but to make evidence retrieval, consolidation, and updating more scalable and traceable (Celli & Mingazov, 2024). A further focus lies on mapping measures to desired effects. AI-Health-Path therefore investigates approaches for linking measures, objectives, effects, and side effects in a more structured way. These mappings must operate within existing legal and regulatory frameworks, including national pandemic policies and international guidance such as WHO and ECDC frameworks. It is important to note that AI-Health-Path examines approaches towards this vision; a fully operational AI-driven policy system does not yet exist. By distinguishing between processes that may be automated, tasks that may be AI-supported, and decisions that must remain human, the study highlights why AI literacy is necessary for responsible hybrid decision-making in health crises.

4. AI-enhanced knowledge management for decision support

AI-enhanced decision support in health crises depends on transforming distributed, heterogeneous, and rapidly changing information into traceable evidence. Knowledge management is therefore not merely a technical backend function, but determines how evidence is retrieved, stored, updated, linked to policy objectives, and made available for human interpretation. For decision-making, it is crucial to provide dynamic, accurate, and context-sensitive knowledge (Rhem, 2020). Three technical developments are especially relevant for such evidence pathways:

First, knowledge graphs and retrieval-augmented generation (RAG) can help ground AI-generated outputs in curated and versioned document repositories. Instead of relying only on the internal parameters of a language model, RAG retrieves information from defined sources before generating a response (Lewis et al., 2020). In health-crisis contexts, this can support traceability because outputs can be linked back to specific documents, such as WHO guidance, ECDC recommendations, or peer-reviewed literature.

Second, AI agents may support defined workflow steps, such as screening sources, flagging updates, comparing evidence, or preparing structured summaries. In high-stakes contexts, however, such agents should be understood as bounded workflow tools rather than autonomous decision actors. Their tasks, data access, and action spaces must be clearly limited, documented, and be subject to human oversight. This is consistent with calls for clearly defined action spaces, oversight checkpoints, and interpretable logs in agentic AI systems (Shavit et al., 2023).

Third, decentralised data management is necessary because relevant health-crisis evidence is distributed across institutions, sectors, and governance levels. Data spaces and federated infrastructures can enable governed access to distributed evidence without requiring all data to be pooled in a single central repository. This is especially relevant where data sovereignty, privacy, institutional control, and legal compliance must be maintained. Data spaces have therefore been described as governance-compatible infrastructures for sovereign data sharing (Otto et al., 2022).

Together, knowledge graphs, RAG-based retrieval, bounded AI agents, and decentralised data infrastructures can support more traceable and accountable AI-enhanced knowledge management. Figure 1 illustrates how these three components interact in a crisis decision-support context.

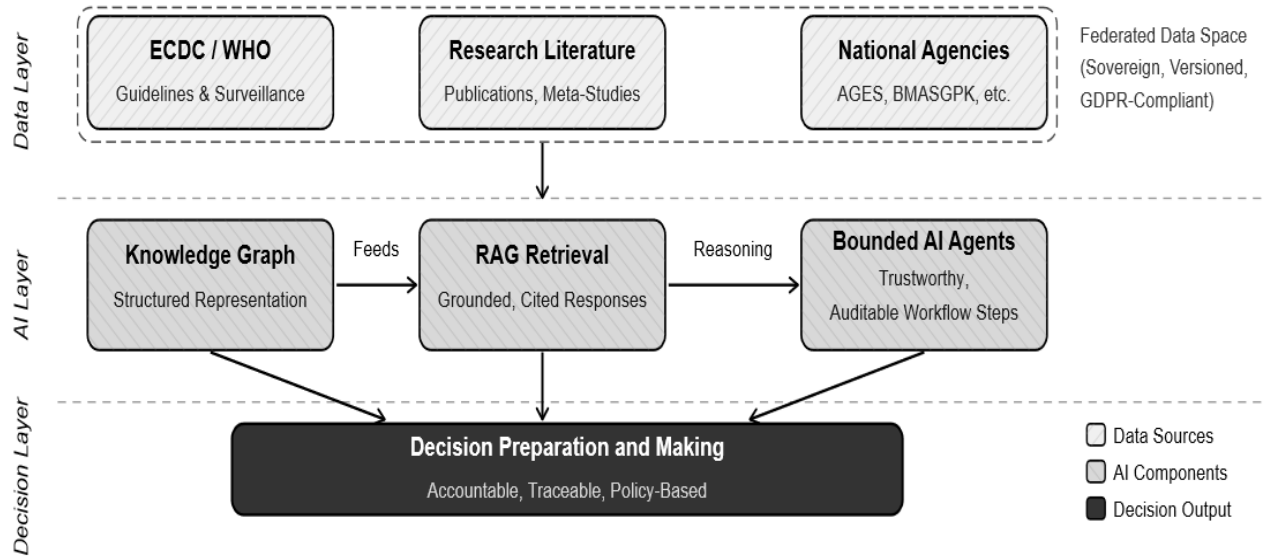


Figure 1. AI-enhanced knowledge management architecture for traceable decision support Source: authors

5. AI literacy and responsible implementation in decision support

5.1. AI literacy as a governance competence

In the context of AI-supported public health policy, AI literacy should be understood less as a narrow technical skill than as a governance-relevant competence that enables human actors to interpret, question, and responsibly use AI-support. Existing AI-literacy frameworks remain heterogeneous, but they converge around a shared core: the ability to understand how AI systems operate, critically evaluate their outputs, recognise limitations and risks, and act responsibly within specific contexts (Annapureddy et al., 2025; Chee et al., 2025; Long & Magerko, 2020).

This is particularly relevant for health-crisis policy-making, where AI may support evidence retrieval, knowledge condensation, risk communication, and structured measure alignment, while final decisions remain embedded in legal, ethical, social, and political responsibility. Empirical and conceptual work suggests that higher AI literacy is associated with trust, perceived fairness, and acceptance of AI systems, while insufficient literacy may contribute to over-trust, misinterpretation, or uncritical reliance on AI-generated outputs (Berman et al., 2024; Scantamburlo et al., 2025; Xu et al., 2025).

Recent public health literature reinforces this point. Health workers are increasingly positioned as intermediaries between algorithmically mediated outputs and patients, communities, and decision-makers, yet AI literacy in this workforce remain underdeveloped and inconsistently measured (Iyamu et al., 2026). Similarly, evidence on AI use in public health decision-making by elected officials remains limited and largely exploratory, indicating that AI adoption in policy settings is still not matched by a mature evidence base on implementation and governance (Campbell et al., 2026). AI-based educational approaches for pandemic preparedness and response show potential for scalable learning, risk communication, and misinformation management, but also highlight gaps in standardised outcome measures, equity analyses, and governance frameworks (Souza et al., 2026).

AI literacy should therefore not be treated as an optional training add-on, but as a structural safeguard across three levels. (1) Policy-makers need sufficient literacy to interpret AI-supported evidence and understand uncertainty. (2) Expert analysts need the competence to validate outputs, detect bias, and document assumptions. (3) Institutions need governance capacity to define responsibilities, ensure transparency, and maintain accountability. This aligns with the EU AI Act, which requires providers and deployers to ensure a sufficient level of AI literacy among staff and others involved in the operation and use of AI system (Regulation (EU) 2024/1689).

In this sense, AI literacy becomes a prerequisite for responsible decision support by enabling the human and institutional oversight needed to ensure transparency and accountability.

5.2. Legal, ethical, and operational risks of AI-enhanced decision support

The use of AI in public health-crisis policy-making entails a range of legal, ethical, and operational risks. These risks arise from factors such as data quality and availability, potential bias, data protection and security gaps, and challenges related to human oversight and institutional capacity (Panteli et al., 2025; Cheong, 2024). Figure 2 provides an overview of the most relevant risks associated with the deployment of AI in this context.

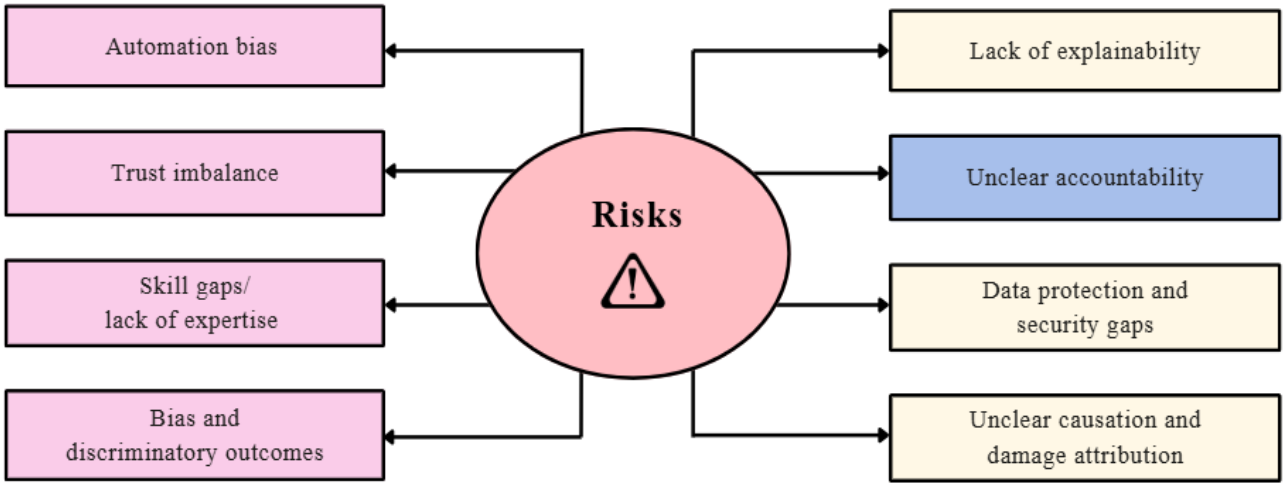


Figure 2. Key risks of AI in decision support Source: authors

Many of the identified risks are not purely technical but stem from limited AI literacy among decision-makers and institutions, affecting their ability to interpret, question, and responsibly act upon AI-generated evidence.

5.3. Safeguards for responsible and trustworthy implementation

To address these risks and ensure that AI systems are deployed responsibly, a comprehensive set of technical, governance, and ethical safeguards is required. These safeguards aim to promote transparency, accountability, fairness, and trust, while ensuring that human decision-makers remain meaningfully involved throughout the policy cycle in health crises (Panteli et al., 2025; Cheong, 2024; Mylrea & Robinson, 2023). Figure 3 summarises the core safeguards that can mitigate identified risks and support responsible AI use in public health emergencies.

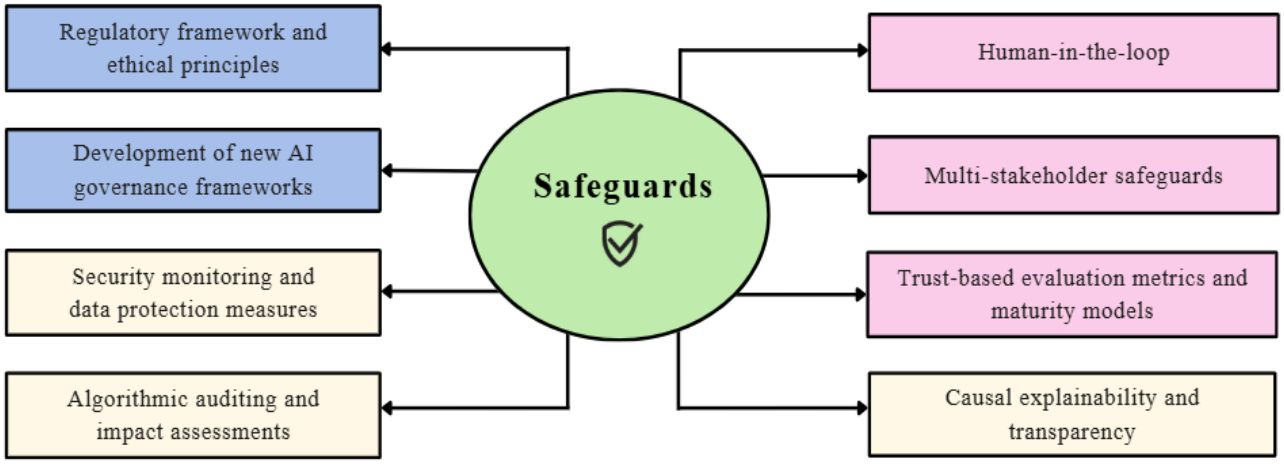


Figure 3. Key safeguards for responsible AI use in decision support Source: authors

For future projects, the consistent implementation of the above-mentioned safeguards is crucial for addressing evolving risks and supporting ethical, transparent, and accountable AI development over time.

5.4. Evidence pathways under human oversight

To support evidence-informed decision-making during health crises, structured pathways are required that integrate AI with human expertise. Figure 4 illustrates an AI-supported evidence pathway for decision support during health crises.

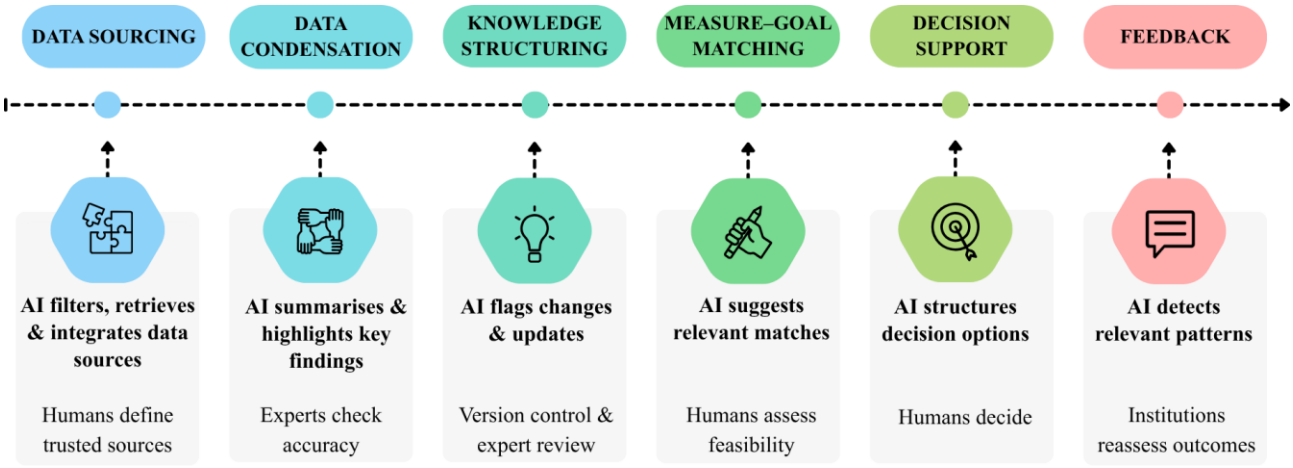


Figure 4. AI-supported evidence pathway for decision support in health crises Source: authors

AI supports data sourcing, evidence condensation, knowledge structuring, and option generation. Human experts are involved at every step and ensure validation, contextual interpretation, and decision-making. Continuous feedback loops enable learning and adaptation across institutional levels. Ultimate decision-making responsibility remains with humans (Lazaros et al., 2026).

6. Opportunities and limits of AI-enhanced decision support

6.1. From manual evidence management to AI-supported evidence pathways

The preceding sections illustrate a shift from structured, expert-driven decision support towards AI-supported evidence formation for health-crisis policy. ROADS to Health showed how pandemic-relevant information, strategic objectives, possible measures, expected effects, and unintended side effects can be organised in a more systematic way (Rainer et al., 2023). Its value does not lay in replacing policy judgement, but in making the relationship between crisis-management targets and possible interventions more explicit. At the same time, ROADS to Health revealed both the potential of structured evidence-based decision support and the limitations of manually maintained approaches. In dynamic health crises, evidence, recommendations, pathogen characteristics, social conditions, and political priorities can change rapidly, making such knowledge bases resource-intensive to maintain, difficult to scale, and vulnerable to delays. This provided the basis for AI-Health-Path's focus on AI-supported data flows and dynamic evidence consolidation.

AI-Health-Path responds to this limitation by investigating how AI can support dynamic evidence formation, automated data consolidation, and structured alignment of measures. Its potential contribution lies in retrieving relevant information, condensing large volumes of evidence, detecting updates, supporting structured comparison, and making relationships between objectives, measures, effects, and side effects more traceable.

In line with the broader conference theme of AI-based decision support and decentralised data management, the infrastructures that enable AI-supported crisis decision-making require further examination. This includes federated data spaces, governed knowledge repositories, and resilient

communication infrastructures. Projects such as AI4Emergency highlight the importance of reliable, low-latency, and scalable communication networks, particularly for applications such as digital situation assessment, voice interaction, and decision-support under high load or partial infrastructure failure (AI4Emergency, 2026).

6.2. Maintaining human accountability in AI-enhanced decision support

This contribution must be understood as decision support, not decision-making. Public health policy decisions are not merely technical optimisation problems; they involve legal constraints, ethical trade-offs, proportionality, public trust, social acceptability, and potential unequal effects across population groups (SAPEA, 2025; WHO, 2021). Even if an AI-supported tool identifies a measure as suitable for a defined objective, human actors must assess whether it is legally permissible, ethically defensible, socially acceptable, and feasible within the given institutional context.

For this reason, responsible AI-supported policy tools require traceability and clearly defined human oversight. Decision-makers and expert analysts must be able to understand which sources informed an output, which assumptions were made, which evidence versions were used, and where uncertainty remains. Technical approaches such as RAG, bounded AI agents, and federated data infrastructures may support this goal by grounding outputs in curated sources, structuring workflow steps, and enabling access to distributed evidence (Lewis et al., 2020; Otto et al., 2022; Shavit et al.,2023). However, these approaches only strengthen policy support if they are embedded in governance structures that ensure auditability, expert validation, legal compliance, and accountability.

6.3. AI literacy and governance as safeguards

AI literacy is therefore a prerequisite for responsible use (Long & Magerko, 2020; Regulation (EU) 2024/1689). Policy-makers need sufficient competence to interpret AI-supported evidence and avoid overreliance on apparently authoritative outputs. Expert analysts need the ability to validate AI-generated summaries, inspect assumptions, detect bias, and document uncertainty. Institutions need governance capacity, including clear responsibilities, validation procedures, documentation standards, and escalation mechanisms. AI literacy should therefore not be understood merely as individual tool competence, but as part of institutional preparedness for AI-supported governance.

At the same time, literacy alone is not sufficient. It must be combined with technical safeguards, transparent data practices, legal and ethical review, and meaningful human control. Overall, AI-Health-Path can contribute to future health-crisis governance by addressing a bottleneck identified through ROADS to Health: the difficulty of maintaining, updating, and applying complex evidence bases during dynamic crises. The challenge is not simply to introduce AI into policy-making, but to design AI-supported evidence pathways that remain accountable to human judgement, public values, and democratic responsibility. If embedded in human-centred governance, transparent data practices, and institutional AI literacy, AI-enhanced knowledge management may support more responsive, evidence-informed, and accountable health-crisis policy. Without these conditions, these systems risk increasing opacity, misplaced trust, and dependence on outputs whose limitations are insufficiently understood.

7. Conclusion and outlook

This paper has argued that AI-enhanced decision support in health-crisis policy should strengthen evidence infrastructure rather than replace human decision-making. The transition from ROADS to Health to AI-Health-Path illustrates how AI may support dynamic evidence formation, automated

data consolidation, and structured alignment of measures. This addresses a practical bottleneck in health-crisis governance: keeping complex evidence bases usable and up to date under rapidly changing conditions. In this context, AI improves the informational conditions under which policy-makers act. AI-supported systems may help retrieve, condense, update, and structure relevant evidence, but they cannot resolve normative trade-offs, determine political priorities, or assume legal and democratic responsibility. Responsible use therefore depends on traceable data sources, auditable workflows, transparent documentation, expert validation, and clearly defined human oversight. AI literacy is a central prerequisite for this process. Policy-makers, expert analysts, and institutions must be able to interpret AI-supported outputs, recognise uncertainty and bias, and govern such systems responsibly.

Taken together, one of the next steps for tackling complex health crises is not autonomous policy-making, but accountable decision-support infrastructure. In line with the broader focus of discovering new paths to tackle complex crises, this includes AI-supported knowledge management, decentralised and traceable data access, resilient communication infrastructures, and institutional AI literacy. These elements are interdependent: AI-based decision support depends on reliable evidence flows, governed data sharing, communicable outputs, and human actors who are able to understand and challenge system results. The key question is therefore not whether AI should be used, but under which organisational, technical, legal, and epistemic conditions it can support responsible human decision-making. This question opens the broader discussion of new paths to tackle complex crises.

Acknowledgement

The research leading to these results received funding from the Austrian security research programme KIRAS, administered by the Austrian Research Promotion Agency (FFG), under ROADS to Health, project number FO999899442, and AI-Health-Path, project number FO999926224.

References

AI4Emergency: KI-optimierte Einsatzkommunikation. (2026). Österreichische Forschungsförderungsgesellschaft (FFG). https://projekte.ffg.at/projekt/5140533

AI-Health-Path: AI-basierte Lösungsansätze zur Entwicklung von dynamischer Evidenzbildung für Entscheidungen von Gesundheitspfaden. (2025). Österreichische Forschungsförderungsgesellschaft (FFG). https://projekte.ffg.at/projekt/5137290

Annapureddy, R., Fornaroli, A., & Gatica-Perez, D. (2024). Generative AI Literacy: Twelve defining competencies. Digital Government Research and Practice, 6(1), 1–21. https://doi.org/10.1145/3685680

Berman, A., De Fine Licht, K., & Carlsson, V. (2024). Trustworthy AI in the public sector: An empirical analysis of a swedish labor market decision-support system. Technology in Society, 76, 102471. https://doi.org/10.1016/j.techsoc.2024.102471

Campbell, E. A., Goodtree, H., Gillani, S., Oyefolu, O., Kelly, A., Rivers, C., & Watson, C. (2026). Impact, use, and implications of artificial intelligence in public health decision making by elected officials: a scoping review. Frontiers in Public Health, 14, 1745684. https://doi.org/10.3389/fpubh.2026.1745684

Celli, F., & Mingazov, D. (2024). Knowledge extraction from LLMs for scalable historical data annotation. Electronics, 13(24), 4990. https://doi.org/10.3390/electronics13244990

Chee, H., Ahn, S., & Lee, J. (2024). A competency framework for AI literacy: Variations by different learner groups and an implied learning pathway. British Journal of Educational Technology, 56(5), 2146–2182. https://doi.org/10.1111/bjet.13556

Cheong, B. C. (2024). Transparency and accountability in AI systems: Safeguarding wellbeing in the age of algorithmic decision-making. Frontiers in Human Dynamics, 6. https://doi.org/10.3389/fhumd.2024.1421273

European Centre for Disease Prevention and Control. (2025). Recommendations for preparedness planning for public health threats. https://www.ecdc.europa.eu/en/publications-data/recommendations-preparedness-planning-public-health-threats

European Parliament and Council of the European Union. (2024). Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). EUR-Lex. https://eur-lex.europa.eu/eli/reg/2024/1689/oj

Iyamu, I., Wheelans, C., Haag, D., Roe, I., Chang, H., Donelle, L., Ellis, U., McKee, G., Bartlett, S., & Gilbert, M. (2026). AI and algorithmic literacy among health workers: A scoping review through a digital health literacy lens. Frontiers in Public Health, 14, 1802392. https://doi.org/10.3389/fpubh.2026.1802392

Lazaros, K., Vrahatis, A. G., & Kotsiantis, S. (2026). Human-in-the-loop artificial intelligence: A systematic review of concepts, methods, and applications. Entropy, 28(4), 377. https://doi.org/10.3390/e28040377

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W., Rocktäschel, T., Riedel, S., & Kiela, D. (2020). Retrieval-Augmented Generation for Knowledge-Intensive NLP tasks. arXiv. https://arxiv.org/abs/2005.11401

Long, D., & Magerko, B. (2020). What is AI literacy? Competencies and design considerations. *Proceedings of the 2020* CHI Conference on Human Factors in Computing Systems. https://doi.org/10.1145/3313831.3376727

Mylrea, M., & Robinson, N. (2023). Artificial intelligence (AI) trust framework and maturity model: Applying an entropy lens to improve security, privacy, and ethical AI. Entropy, 25(10), 1429. https://doi.org/10.3390/e25101429

Otto, B., Hompel, M. T., & Wrobel, S. (2022). Designing data spaces. https://doi.org/10.1007/978-3-030-93975-5

Panteli, D., Adib, K., Buttigieg, S., Goiana-Da-Silva, F., Ladewig, K., Azzopardi-Muscat, N., Figueras, J., Novillo-Ortiz, D., & McKee, M. (2025). Artificial intelligence in public health: Promises, challenges, and an agenda for policy makers and public health institutions. The Lancet Public Health, 10(5), e428–e432. https://doi.org/10.1016/s2468-2667(25)00036-2

Rainer, K., Grosser, C., Hagendorn, M., Leidwein, A., Neubauer, G., Sturm, N., Geyer, C., Salomon, G., Prüller, R., & Schöggl, F. (2023). From strategic planning to tackling pandemics. In P. Doucek, M. Sonntag, & L. Nedomova (Eds.), IDIMT-2023: New challenges for ICT and management: 31st Interdisciplinary Information Management Talks, Hradec Králové, Czech Republic (pp. 29–38). Trauner Verlag. https://doi.org/10.35011/IDIMT-2023-29

Rhem, A. J. (2020). AI ethics and its impact on knowledge management. AI And Ethics, 1(1), 33–37. https://doi.org/10.1007/s43681-020-00015-2

ROADS to Health: Entscheidungsunterstützung für das Pandemiemanagement der Zukunft. (2023). Österreichische Forschungsförderungsgesellschaft (FFG). Retrieved May 29, 2026, from https://projekte.ffg.at/projekt/4658055

SAPEA (2025). Artificial Intelligence in Emergency and Crisis Management: Rapid Evidence Review Report. Munich: SAPEA. https://doi.org/10.5281/zenodo.17737962

Scantamburlo, T., Cortés, A., Foffano, F., Barrué, C., Distefano, V., Pham, L., & Fabris, A. (2024). Artificial Intelligence across Europe: A study on awareness, attitude and trust. IEEE Transactions on Artificial Intelligence, 6(2), 477–490. https://doi.org/10.1109/tai.2024.3461633

Shavit, Y., Agarwal, S., Brundage, M., Adler, S., Campbell, R., Lee, T., Mishkin, P., Eloundou, T., Hickey, A., Slama, K., Ahmad, L., McMillan, P., Beutel, A., Passos, A., Robinson, D. G., & O'Keefe, C. (2023). Practices for governing agentic AI systems. https://cdn.openai.com/papers/practices-for-governing-agentic-ai-systems.pdf

Souza, E. C. S. G., Santos, A. G. D., Junior, Félix, A. M. S., Borges, J. P. A., De Oliveira, L. B., Carneiro, L. M., & De Sousa, A. F. L. (2026). Use of artificial intelligence in public health education for pandemic preparedness and response. Annals of Global Health, 92(1), 21. https://doi.org/10.5334/aogh.5130

World Health Organization. (2021). Ethics and governance of artificial intelligence for health. https://www.who.int/publications/i/item/9789240029200.

World Health Organization. (2024). Global guidance on monitoring public health and social measures policies during health emergencies. https://www.who.int/publications/i/item/9789240094444

Xu, Y., Chen, Z., & Dong, M. (2025). Shaping the fairness journey: The roles of AI literacy, explanation, and interpersonal interaction in AI interviews. International Journal of Human-Computer Studies, 205, 103629. https://doi.org/10.1016/j.ijhcs.2025.103629

CHANGE MANAGEMENT MODELS IN THE AGE OF INDUSTRY 4.0: ARE TRADITIONAL APPROACHES STILL RELEVANT AND IN USE?

Lukáš Klarner, Petr Řehoř, Jaroslav Vrchota, Petra Matoušková, Kamila Němečková, Monika Maříková, Ladislav Rolínek

University of South Bohemia in Ceské Budejovice
Faculty of Economics
klarnl00@ef.jcu.cz, rehor@ef.jcu.cz, vrchota@ef.jcu.cz, pmatouskova@jcu.cz, nemeck07@ef.jcu.cz, marikova@ef.jcu.cz, rolinek@ef.jcu.cz

DOI: 10.35011/IDIMT-2026-25

Keywords

Industry 4.0; change; management of change; artificial intelligence; SMEs.

Abstract

Implementing organizational changes is becoming an increasingly critical discipline for businesses as a wide range of technologies rises. The question, however, is whether businesses actually follow traditional, proven concepts, like Lewin's and Kotter's models, or other concepts offered by literature. In particular, the dynamic environment of Industry 4.0 raises the question of whether these traditional methods are being used. The main aim of this paper is to empirically verify whether enterprises in the Czech Republic use traditional change management models and whether their use is associated with successful change implementation and Industry 4.0 adoption. Based on a sample of 224 companies from a questionnaire survey, the paper examines the use of classical change models in the implementation of changes and modern technologies. Correlation analysis and the Mann-Whitney U test are used to verify these relationships. The results indicate very weak relationships between change models and change implementation, and these relationships are not statistically significant. While change models do show statistical significance in relation to Industry 4.0, these differences are very weak. The study contributes to the ongoing discussion on the relevance of traditional change management models in Industry 4.0 environments.

1. Introduction

In the fast-paced world of the 21st century, businesses need to achieve their desired outcomes as quickly and effectively as possible. Given that this era is marked by constant change, mastering the discipline that addresses these changes – change management – is essential to achieving those outcomes (Guo & Polak, 2023). This field is not limited to organizational change or development; on the contrary, it is currently also closely linked to innovation, technology, and digital transformation (Abu Orabi et al., 2024). Naturally, even today, organizational changes require consideration of several factors, such as corporate culture, the company's vision and mission, strategy, and stakeholders (Phillips & Klein, 2023).

Change management and digital transformation (Industry 4.0) influence each other. Properly implemented change management is key to the successful implementation of Industry 4.0 in organizations, and it is advisable to follow best practices (Brodeur et al., 2023); however, digital transformation also has great potential to transform change management processes and frameworks (Thakur et al., 2025). In this field, AI can help overcome employee resistance and facilitate the deeper implementation of Industry 4.0, for example, through predictive analytics (Jagtap & Suvarna, 2025).

The current debate highlights the need to find alternative methods to the traditional approach to change management (Lies, 2020). Moreover, this approach has several limitations, including the inability to identify positive links between the organization's external and internal environments (Sopow, 2020) and a significant decline in employee performance during periods of ongoing change (Panagiotopoulos & Chatzis, 2025). For this reason, we are seeing modern approaches to change management and the development of new frameworks (Riaz et al., 2023).

Nevertheless, traditional change management techniques offer high-quality, practical principles for managing organizational change. Among the best known are Lewin's and Kotter's models (Harrison et al., 2021); however, the literature also offers a wide range of others, such as McKinsey 7S, ADKAR, Satiro's growth model, Nudge theory, and Bridges' transition model (Kamarova et al., 2024). The question, then, is whether traditional approaches to change management are still valid in the era of digital transformation (Kherrazi & Roquilly, 2025). This study provides empirical evidence on the actual use and relevance of traditional models in SMEs undergoing digital transformation.

2. Methodology

The main aim of this paper is to empirically verify whether enterprises in the Czech Republic use traditional change management models and whether their use is associated with successful change implementation and Industry 4.0 adoption. To achieve the main aim, data were collected through a questionnaire survey, with 224 companies participating. The questionnaire was available and completed exclusively online. The respondents were company managers and owners. Respondents were contacted exclusively via email sent to their work email addresses. Their email addresses were obtained from a university database. Data collection began in late summer 2024 and lasted exactly 136 days. Before the empirical study began, a pilot test of the questionnaire was conducted on a sample of 10 respondents. The main aim of this pilot test was to verify the questionnaire's validity. Based on the responses received, minor revisions were applied. The respondents to this pilot survey also included business owners and managers.

The research instrument was developed in accordance with recommendations from the professional literature on questionnaire design (Plengkham et al., 2025). The questionnaire was divided thematically into three interconnected sections. The first section consisted of identification questions (6 questions in total), followed by questions on change management (11 questions), and then questions on Industry 4.0 and modern technologies (17 questions). A combination of questions was used to elicit both quantitative and qualitative data. A total of 8 open-ended questions and 26 closed-ended questions were used. For the purpose of this paper 4 questions were used. Among the closed-ended questions, dichotomous and Likert-scale questions were used, with a 1–4 scale. The even number of options on the scale prevents respondents from selecting a neutral choice, which is often referred to as the "socially desirable" answer. At the same time, however, an even-numbered scale is less common in social science research (Kusmaryono et al., 2022).

Non-probability sampling was used to collect the data, specifically quota sampling (Futri et al., 2022). Quotas were set based on the population according to CZ-NACE, and respondents were approached to ensure the sample reflected the actual sectoral structure. This ensured that companies from all

sectors contributed to the research, as change management and Industry 4.0 affect all companies in the economy. During the research, the basic ethical principles were followed: participation was voluntary, and respondents were informed of the research's purpose and how the data would be handled. At the same time, they were informed that all data would be presented in aggregated form.

Based on the literature review, the main research question (RQ) of this study was formulated as follows: "*Do enterprises using traditional change management models achieve more successful change implementation, and how does this relationship vary with Industry 4.0 technologies?"* Two hypotheses are derived from the established RQ.

- H_1: Enterprises using traditional models have a higher success rate in implementing change.
- H_2: Industry 4.0 positively influences change management in the form of higher change success rates.

The following traditional models of organizational change management were examined: Lewin's, Kotter's, ADKAR, McKinsey 7S, PDCA, Bridges' transition model, Nudge theory, Satiro's model, and the Kübler-Ross curve. The models analysed were selected because they are among the most frequently cited and widely used traditional approaches in the change management literature. Given that the proportion of responses in individual categories regarding the practical use of change models was very low in some cases, we adjusted the data distribution (Iacobucci et al., 2025) by merging two adjacent categories. The result is that only two categories remain: "we use" and "we do not use." The success of change implementation was measured using respondents' self-assessments on a Likert scale as part of the questionnaire, as well as the degree of use of Industry 4.0 technologies. However, the positive impact of change management was not measured; rather, the relationship between Industry 4.0 and the success of the changes was observed.

Empirical studies often evaluate the use of traditional concepts and models of change management and the subsequent assessment of the success of change implementation (Sujová & Simanová, 2022). There is also ongoing debate about Industry 4.0 and its impact on change management (Cimini et al., 2020). Therefore, the hypotheses formulated here expand and complement the existing body of knowledge in this field of research. Spearman's rank correlation (Stephanou & Varughese, 2021) and the Mann-Whitney U test (Chechile, 2020) were used to evaluate the hypotheses. The Mann-Whitney U test was chosen because it compares an ordinal variable between two independent groups. Before applying the test, the independence of the groups and the ordinal nature of the data were verified. However, descriptive statistics and basic measures of central tendency – specifically the weighted average – were also an equally important part of the methods employed. The number of employees is used as the weight in the calculation of the average. All calculations were performed at the 95% significance level and carried out in RStudio.

3. Results

The introduction provides an overview of the research sample's structure. As mentioned above, 224 respondents participated in the study. In terms of size structure, as defined by European Union recommendations (Pedraza, 2021), there are 110 micro-enterprises, 71 small enterprises, and 43 medium-sized enterprises. The average number of employees in the sample is 28.69. In terms of sectoral structure, the strongest representation comes from manufacturing (78 enterprises), construction (22 respondents), wholesale/retail trade (25 enterprises), agriculture (16 organizations), and financial and insurance activities (13 enterprises); however, all sectors are represented.

On the scale used, a lower value always indicates a higher level of model use, change success, or Industry 4.0 adoption. As mentioned in the Methodology chapter, the individual change models were merged into categories 1 and 2, as well as 3 and 4.

The weighted average score for the variable related to successful change implementation is 2.013, indicating that organizations are satisfied with their change management, as the changes introduced are generally implemented successfully. Regarding the implementation of Industry 4.0 and modern technologies, the weighted score is 3.064, indicating significant room for further and deeper implementation. From a descriptive perspective, it is evident that companies do not actually use the vast majority of change models in practice. Table 1 presents the weighted averages for all the models examined. The results clearly indicate that the actual use of traditional approaches in practice is significantly limited, and companies rarely use these models.

Table 1. Weighted average of usage of change management models

Lewin	Kotter	ADKAR	Bridges	PDCA	7S	Nudge	Satir	Kübler-Ross
3.218	3.389	3.469	3.756	3.557	3.499	3.924	3.827	3.825

Source: authors

Subsequently, statistical tests were used to determine whether the implementation of changes was related to the use of change models. First, a correlation analysis was performed to examine the correlation between the implementation of changes and each of the traditional models separately. The resulting coefficients are presented in Table 2. If a value is marked with an asterisk (*), the coefficient is statistically significant (p-value 0.05).

Table 2. Correlation coefficients of change models in relation to the implementation of changes

Lewin	Kotter	ADKAR	Bridges	PDCA	7S	Nudge	Satir	Kübler-Ross
-0.180*	-0.195*	-0.126	-0.079	-0.083	-0.081	0.002	-0.055	0.003

Source: authors

As Table 2 shows, none of the traditional change models exhibits a strong correlation with successful change implementation. However, negative coefficients must be interpreted with caution in this case, since the higher the value for the use of change models (0 or 1), the lower the value for the implementation of changes – where, however, low values indicate high success in their implementation. Although several relationships are statistically significant, the observed effect sizes and correlation coefficients are very small. The practical impact of traditional models on the implementation of changes, therefore, appears to be limited. For verification purposes, an average was also calculated for all the traditional models under consideration and subjected to a correlation analysis. Although the correlation coefficient found is statistically significant (p-value < 0.05), its value of -0.187 does not indicate that the use of models influences the implementation of changes.

This was followed by an examination of whether there were any direct differences between the groups in model use (according to change implementation), with one group answering “yes, we use them” and the other “no, we do not use them.” Given the ordinal and binary nature of the data, the Mann-Whitney U test was used. In addition to p-values, the W test statistic and the effect size were also calculated. Results are provided in Table 3. First and foremost, it is important to note that none of the observed effect sizes are large. The test identified two traditional change models that are statistically significant: Lewin’s (p-value = 0.007) and Kotter’s (p-value = 0.004). It is evident that, despite the observed statistical significance in the differences between groups, the practical real-world impact is very weak, as indicated by the very low effect size and the very low correlation coefficient.

At this point, the first hypothesis can be addressed. At the 95% level, it was not demonstrated that companies using traditional models have a higher success rate in implementing. However, the results

must be interpreted with caution, as models may still offer certain advantages in practice. Based on the findings, however, these advantages are likely to be quite limited. It should also be noted that some companies may use models only partially or only in selected situations.

Table 3. Mann-Whitney U Test results – implementation of changes

Model	p-value	W statistics	Effect size
Lewin	0.007	4,496	0.180
Kotter	0.004	2,975	0.194
ADKAR	0.061	2,608	0.126
Bridges	0.242	819	0.079
PDCA	0.219	1,514	0.082
7S	0.229	1,617	0.081
Nudge	0.979	110	0.002
Satir	0.412	414	0.055
Kübler-Ross	0.966	218	0.003

Source: authors

Next, we examined the relationships between change management and Industry 4.0. First, we calculated the correlation coefficient between the success of change implementation and the degree of Industry 4.0 adoption in companies. This coefficient is 0.147, with a p-value of 0.027, indicating statistical significance. A weak association was observed between Industry 4.0 adoption and reported change implementation success. Subsequently, correlations among individual models regarding the degree of Industry 4.0 adoption were examined. The results are presented in Table 4. Where an asterisk (*) appears, the coefficient is statistically significant.

Table 4. Correlation coefficients of change models in relation to the degree of Industry 4.0 adoption

Lewin	Kotter	ADKAR	Bridges	PDCA	7S	Nudge	Satir	Kübler-Ross
-0.234*	-0.153*	-0.123	-0.068	-0.143*	-0.169*	-0.033	-0.102	-0.101

Source: authors

The correlation coefficients found are negative in all cases, indicating an inverse relationship. However, this inverse relationship must be interpreted with caution, just as in the case of Table 2.

This was followed by an examination of whether there were any direct differences between the groups in model use (according to Industry 4.0 use), with one group answering “yes, we use them” and the other “no, we do not use them.” Given the ordinal and binary nature of the data, the Mann-Whitney U test was used. In addition to p-values, the W test statistic and the effect size were also calculated. Results are provided in Table 5. In this case, the Mann-Whitney U test revealed several statistically significant differences. However, the observed effect sizes are small to moderate, so the results must be interpreted with caution. The small effect size corresponds to the low correlation coefficients.

It should be noted, however, that the use of specific traditional change models in relation to the proportion of activities implemented with Industry 4.0 shows several significant differences. At the aggregate level, however, no influence of Industry 4.0 on the implementation of changes was demonstrated within the correlation. It is therefore possible to address the second hypothesis under investigation, namely that, at the 95% level, Industry 4.0 does not positively influence change management, as evidenced by higher success rates in implementing changes.

Table 5. Mann-Whitney U Test results – Industry 4.0 use

Model	p-value	W statistics	Effect size
Lewin	<0.001	4,776	0.233
Kotter	0.022	2,822	0.153
ADKAR	0.066	2,605	0.123
Bridges	0.314	798	0.068
PDCA	0.033	1,695	0.142
7S	0.012	1,891	0.168
Nudge	0.628	141	0.033
Satir	0.129	486	0.102
Kübler-Ross	0.132	348	0.101

Source: authors

4. Discussion

The findings of this study suggest that the actual use of traditional change management models in Czech small and medium-sized enterprises is rather limited and that their impact on the success of change implementation is weak. These results are partly consistent with the study by Sujová & Simanová (2022), which also found a relatively low rate of formal model use in practice. However, unlike their findings, which suggest a more significant relationship between the functional approach and successful change outcomes, our results show only very weak correlations and negligible effect sizes. This discrepancy can be explained by differences in national context, industry composition, or methodological approaches, particularly the aggregation of models into binary categories.

Furthermore, the limited impact of traditional models observed in this research corresponds with the arguments presented by Lies (2020) and Riaz et al. (2023), who emphasize the growing inadequacy of traditional linear approaches in dynamic and complex environments. The weak relationship between model usage and change success found in this study supports the idea that organizations increasingly rely on more flexible, adaptive, or informal approaches rather than strictly following predefined frameworks. On the other hand, the moderate statistical significance observed for Lewin's and Kotter's models suggests that these classic frameworks may still retain some relevance, particularly as guidelines rather than rigid methodologies.

In the context of Industry 4.0, the results show only a very weak positive relationship between digital transformation and change success, which contrasts with expectations in the literature. For example, Cimini et al. (2020) emphasize a stronger positive influence of Industry 4.0 technologies on organizational change processes. Similarly, Westover (2024) highlights the transformative potential of artificial intelligence in improving change management outcomes. The discrepancy with our findings may suggest that Czech small and medium-sized enterprises have not yet fully exploited the potential of Industry 4.0 technologies, or that more time, resources, or further organizational changes are needed to realize their benefits fully.

5. Conclusion

The main aim of this paper was to empirically verify whether enterprises in the Czech Republic use traditional change management models and whether their use is associated with successful change implementation and Industry 4.0 adoption. The results suggest that although some traditional models

show statistically significant relationships, their overall practical impact is very weak. At the same time, Industry 4.0 technologies show only a minimal positive correlation with the success of change. Overall, the results suggest that neither traditional models nor digital transformation alone is sufficient to determine change implementation in SMEs. Neither H_1 nor H_2 was supported.

Several factors may account for the observed lack of integration. Organizations may perceive these frameworks as overly formalized or better suited to large corporations. Furthermore, many firms may prefer flexible, experience-based approaches that are not explicitly linked to established change management theories. From a managerial perspective, the findings suggest that traditional change management models should be viewed as supporting rather than decisive tools in digital transformation processes. From a theoretical point of view, the findings extend current knowledge on the applicability of traditional change management frameworks within digitally transforming organizations.

However, the study has several limitations that must be acknowledged. First, the use of non-probability quota sampling limits the generalizability of the results. Second, relying on self-reported data may lead to subjective bias. Third, simplifying the application of models into binary categories may have reduced the ability to capture subtle differences in implementation. Another limitation is that respondents, particularly those from micro-enterprises, may not have been familiar with the technical terminology associated with change management models. As a result, some organizations may apply principles similar to those of traditional models without explicitly recognizing or naming them. Last but not least, one limitation of the study is that it did not examine what alternative approaches companies use in place of traditional models. Future research should therefore focus on more detailed measurements of model implementation, set new research design to capture causal relationships, and cross-country comparative studies. Furthermore, future research could explore hybrid or adaptive approaches to change management that better reflect the reality of Industry 4.0.

Acknowledgement

This paper was supported by Grant Agency of University of South Bohemia in České Budějovice GA JU 068/2024/S.

References

Abu Orabi, T., Abu Alfalayeh, G., Alhyasat, W. B. A. K., Ababne, A., Alkhawaldah, R., & Qteishat, M. (2024). Change management in business organization: A literature review. Human Systems Management, 43(2), 195-213. https://doi.org/10.3233/hsm-230031

Brodeur, J., Deschamps, I., & Pellerin, R. (2023). Organizational changes approaches to facilitate the management of Industry 4.0 transformation in manufacturing SMEs. Journal of Manufacturing Technology Management, 34(7), 1098-1119. https://doi.org/10.1108/jmtm-10-2022-0359

Chechile, R. A. (2020). A Bayesian analysis for the Mann-Whitney statistic. Communications in Statistics - Theory and Methods, 49(3), 670-696. https://doi.org/10.1080/03610926.2018.1549247

Cimini, C., Boffelli, A., Lagorio, A., Kalchschmidt, M., & Pinto, R. (2020). How do industry 4.0 technologies influence organisational change? An empirical analysis of Italian SMEs. Journal of Manufacturing Technology Management, 32(3), 695-721. https://doi.org/10.1108/jmtm-04-2019-0135

Futri, I. N., Risfandy, T., & Ibrahim, M. H. (2022). Quota sampling method in online household surveys. MethodsX, 9, 101877. https://doi.org/10.1016/j.mex.2022.101877

Guo, H., & Polak, P. (2023). Intelligent finance and change management implications. Humanities and Social Sciences Communications, 10(1), 413. https://doi.org/10.1057/s41599-023-01923-4

Harrison, R., Fischer, S., Walpola, R. L., Chauhan, A., Babalola, T., Mears, S., & Le-Dao, H. (2021). Where Do Models for Change Management, Improvement and Implementation Meet? A Systematic Review of the Applications of Change Management Models in Healthcare. Journal of Healthcare Leadership, 13, 85-108. https://doi.org/10.2147/jhl.s289176

Iacobucci, D., Román, S., Moon, S., & Rouziès, D. (2025). A Tutorial on What to Do With Skewness, Kurtosis, and Outliers: New Insights to Help Scholars Conduct and Defend Their Research. Psychology & Marketing, 42(5), 1215-1431. https://doi.org/10.1002/mar.22187

Jagtap, S., & Suvarna, S. R. (2025). AI-driven change management: A secondary research on overcoming employee resistance in digital transformation. International Journal of Financial Management and Economics, 8(2), 963-965. https://doi.org/10.33545/26179210.2025.v8.i2.625

Kamarova, S., Gagné, M., Holtrop, D., & Dunlop, P. D. (2024). Integrating behavior and organizational change literatures to uncover crucial psychological mechanisms underlying the adoption and maintenance of organizational change. Journal of Organizational Behavior, 46(2), 263-287. https://doi.org/10.1002/job.2832

Kherrazi, S., & Roquilly, C. (2025). Managing Digital Transformation in Contemporary Contexts: Are Classical Change Management Frameworks Still Relevant? Journal of Change Management, 25(2), 85-115. https://doi.org/10.1080/14697017.2025.2494508

Kusmaryono, I., Wijayanti, D., & Maharani, H. R. (2022). Number of Response Options, Reliability, Validity, and Potential Bias in the Use of the Likert Scale Education and Social Science Research: A Literature Review. International Journal of Educational Methodology, 8(4), 625-637. https://doi.org/10.12973/ijem.8.4.625

Lies, J. (2020). Imagining Niklas Luhmann as a “change manager”: scrum as an example of a closed system for the self-organized stabilization of corporate agility. Journal of Organizational Change Management, 33(7), 1417-1429. https://doi.org/10.1108/jocm-11-2019-0355

Panagiotopoulos, F., & Chatzis, V. (2025). A Continuous Monitoring and Reallocation Method for Successful Decisions in Change Management. Algorithms, 18(4), 184. https://doi.org/10.3390/a18040184

Pedraza, J. M. (2021). The Micro, Small, and Medium-Sized Enterprises and Its Role in the Economic Development of a Country. Business and Management Research, 10(1), 33-44. https://doi.org/10.5430/bmr.v10n1p33

Phillips, J., & Klein, J. D. (2023). Change Management: From Theory to Practice. TechTrends, 67(1), 189-197. https://doi.org/10.1007/s11528-022-00775-0

Plengkham, B., Rattanasak, S., & Sukserm, P. (2025). 6Ps Guidelines for Constructing an English Questionnaire for a Survey in Social Science Research. Journal of Education and Learning, 14(3), 212-220. https://doi.org/10.5539/jel.v14n3p212

Riaz, S., Morgan, D., & Kimberley, N. (2023). Managing organizational transformation (OT) using complex adaptive system (CAS) framework: future lines of inquiry. Journal of Organizational Change Management, 36(3), 493-513. https://doi.org/10.1108/jocm-08-2022-0241

Sopow, E. (2020). Aligning workplace wellness with global change: an integrated model. Journal of Organizational Change Management, 33(5), 909-923. https://doi.org/10.1108/jocm-11-2019-0334

Stephanou, M., & Varughese, M. (2021). Sequential estimation of Spearman rank correlation using Hermite series estimators. Journal of Multivariate Analysis, 186, 104783. https://doi.org/10.1016/j.jmva.2021.104783

Sujová, A., & Simanová, Ľ. (2022). Management models of changes – the empirical study in Slovak companies. Business, Management and Economics Engineering, 20(01), 23-40. https://doi.org/10.3846/bmee.2022.15397

Thakur, S., Misra, P., & Kumar, A. (2025). The symbiotic relationship: Navigating change management in the era of digital transformation. International Journal of Financial Management and Economics, 8(2), 938-941. https://doi.org/10.33545/26179210.2025.v8.i2.622

Westover, J. (2024). Embracing AI in Change Management. Human Capital Leadership Review, 13(1). https://doi.org/10.70175/hclreview.2020.13.1.11

DECISION SUPPORT FOR THE MONITORING AND MANAGEMENT OF BATHING WATER QUALITY

Georg Neubauer, Talha Anwar, Martin Jung, Verena Parzer

AIT Austrian Institute of Technology GmbH
georg.neubauer@ait.ac.at, talha.anwar@ait.ac.at, martin.jung@ait.ac.ar, verena.parzer@ait.ac.at

Bernd Reichl

AGES Austrian Agency for Health and Food Safety
bernd.reichl@ages.at

DOI: 10.35011/IDIMT-2026-33

Keywords

Decision Support System; Cyanobacteria; Bathing Water Quality; Algal Bloom.

Abstract

In this paper, we focus on the risk assessment of cyanobacteria in bathing water in Austria. To this end, we have developed a modular demonstrator of a decision support system, which provides a structured platform for integrating monitoring data, expert evaluations, and legal thresholds. The system supports consistent, transparent, and timely decision-making and interaction in the management of bathing water quality in Austria. Legal provisions form the basis for the operation of a decision support system designed to assist authorities and other stakeholders in the systematic evaluation, documentation, and management of cyanobacterial events.

1. Introduction

Water is the most important resource for human health and social development. However, climate change along with escalating human activities is putting pressure on this resource. Rising temperatures in surface water and groundwater are causing a deterioration in quality and so playing a significant role in the widespread proliferation of microorganisms in surface waters across many European countries (Raeder, 2022; Istvánovics et al., 2022). Within this paper, the focus is on cyanobacteria (CB), particularly in the context of their mass proliferation (CHAB, cyanobacterial harmful algal bloom), which impairs water quality and limits its suitability for leisure, recreation (Valeriani et al., 2024) and the supply of drinking water. This indirectly results in economic losses, for example in the tourism or fisheries industries (Anderson et al., 2000). Some of these CBs produce toxins (cyanotoxins [CT]) that are harmful to health with effects ranging from skin reactions, inflammation, vomiting, to diarrhoea, liver damage or even paralysis. As temperatures continue to rise, cyanobacterial proliferation is expected to become an increasingly significant challenge across Europe, including Austria. Health incidents involving cyanotoxins were reported from 8 European

countries (the Czech Republic, Estonia, Hungary, Poland, Portugal, Serbia, Sweden and the United Kingdom) including skin and mucosal membrane irritation, fever, gastrointestinal illness, respiratory distress and pulmonary consolidation (Codd, G A. et al., 2005). The Cyano+ project primarily focuses on surface waters designated as bathing waters, particularly those classified as EU bathing sites. These sites are subject to five inspections during the bathing season, in accordance with the Bathing Water Directives (Directive 2006/7/EC; BGBl. II 349, 2009). This sampling approach is used to assess bathing water quality. However, due to the small number of samples and sampling points, its interpretative value is significantly limited. Especially, when considering that cyanobacteria are only usually tested for if the bathing water profile indicates potential for mass proliferation (BGBl. II 349, 2009). As the periods between the onset and disappearance of a CHABs are often brief, they can usually only be observed by chance or through reports from attentive bathing site operators. However, they can easily be mistaken for mass blooms of other algae species. Further questions will be addressed in brief in the frame of Cyano+:

- How do CTs enter the subsoil or groundwater?
- How do they behave in the saturated zone of the subsoil?
- Can this information be used to derive a risk assessment for spring and groundwater-based drinking water abstraction points?

The overall aim is to provide decision support and enable timely, time-resolved assessments of trends and risk levels by developing methods for risk assessment and forecasting, as well as a fluorescence-based sensor that integrates satellite based earth observation, historical weather data, and climate projections and to enable stakeholders, including operators of natural bathing facilities and blue-green infrastructure, public authorities and drinking water suppliers, to take appropriate measures in good time to ensure the sustainable protection of the population, the affected water bodies and drinking water infrastructure.

2. Overview on existing Decision Support Systems

In recent years, Decision Support Systems (DSS) for bathing water quality management have evolved from manual sample-based validation to integrated platforms. The design and functionality of these platforms such as real-time sensing, predictive modelling, or automated notifications are primarily driven by the mandates of the stakeholders using them.

The main purposes of DSS include:

- Legal compliance validation e.g., classifying water status as excellent, good, sufficient, or poor, as required by Directive 2006/7/EC,
- Short-term risk prevention and forecasting of water quality deterioration, often using proxy variables like rainfall, river flow, or temperature,
- Large-scale monitoring across multiple bathing sites.

Early warning DSS focuses on predicting short-term water quality deterioration due to episodic pollution. These systems use proxy variables and can operate independently of, or in combination with, microbiological and chemical analyses. Examples are SWIM:AI and Digital Water.City, which provide near real-time alerts for both authorities and the public (SWIM:AI 2026; DigitalWater.City 2026).

Forecasting systems employ hydrodynamic-water quality models, statistical regression, or data-driven machine learning to predict future conditions using historical and real-time data. These

systems enable proactive management, such as anticipating increased bacterial levels following heavy rainfall (Sene 2024). Predictive modelling simulates complex interactions affecting long-term water quality, considering factors like climate change and land use.

Some DSS integrate diverse data sources such as satellite earth observation, in situ sensors, microbiological sampling, and numerical models. Examples include iSWIM, a web-based DSS for the Black Sea combining numerical models, in situ data, and remote sensing (iSWIM 2026), and AlgaRisk, a satellite-supported monitoring service for the UK Environment Agency (Shutler, J.D. et al (2025)). The Cyano+ DSS demonstrator follows a similar data-integrating approach (see Section 3.3 and 3.4). Systems like the Cyano+ DSS integrating multiple data sources and combining different functional elements are superior to monolithic systems, provided they are built in a modular, adaptable way.

3. Concept of the Cyano+ Decision Support System

3.1. Mandates and actors

The monitoring and management of bathing water quality in Austria is governed by a specifically defined legal and institutional framework that ensures the protection of human health and the early identification of potential risks, including those posed by cyanobacteria and their toxins. At the European level, Directive 2006/7/EC concerning the management of bathing water quality establishes the overall requirements for monitoring, assessment, taking measures and public information. This directive is transposed into national law through the Austrian Bathing Water Hygiene Act (BGBl. I, 249, 1976) and further specified by the Bathing Water Ordinance (BGBl. II 349, 2009), which together define the responsibilities, monitoring parameters, and procedural obligations for competent authorities.

A key responsibility arising from this framework is the regular monitoring of bathing waters, including visual inspections and analytical assessments. In cases of potential health risks due to the presence of cyanobacteria appropriate management measures need to be taken, including the timely information of the public. The implementation of these mandates involves several key actors. Scientific expertise is provided by hygiene experts from the Austrian Agency for Health and Food Safety (AGES), the municipal department MA 39 of the City of Vienna, and the provincial health laboratories of Carinthia and Vorarlberg. These actors are responsible for sampling, laboratory analysis, ecological assessment, and expert interpretation of findings, particularly regarding cyanobacterial abundance and potential toxin production. The district administrative authorities act as the legally competent bodies for decision-making and enforcement. Based on expert assessments, they are responsible for ordering management measures such as bathing warnings, temporary bathing prohibitions, or other risk mitigation actions. Effective cooperation and information exchange between scientific institutions and administrative authorities are therefore essential.

3.2. Workflow – Benchmarking the baseline

The monitoring of bathing water quality is based on various types of data. These include the coordinates of sampling sites, detailed bathing water profiles, and measurement results covering routine microbiological and chemical parameters as well as partially specific data on cyanobacteria, and cyanotoxins. However, data management poses significant challenges: there is no central, dedicated database specifically for cyanobacteria. Instead, data is scattered across different LIMS (Laboratory Information Management System) systems operated by the key actors mentioned above,

reports, tables, and lists. Bathing water profiles are often unstructured and only partially machine-readable.

The current process is characterized by heterogeneous data formats and a lack of consistent linkage between bathing sites, measurements, assessments, and actions. These weaknesses significantly hinder automated processing and the use of data for process support. The standard process for assessing bathing water quality involves several steps: first, on-site visual inspection, sampling and analytical assessments are conducted, followed by the creation of an independent test report. This report is then submitted to the district administrative authority in charge, which decides on any necessary measures. The hygiene experts acting on behalf of the district authorities to monitor the bathing waters are responsible for technical evaluations and for providing recommendations, while the formal decision rests with the authorities. The separation between technical assessment and formal decision-making is a key principle. If limit values are exceeded, warnings, bathing prohibitions, or restrictions may be imposed. These threshold-based measures are time-critical. The reopening of a bathing area is not automatic but requires a separate process. Typical criteria for reopening include evidence that the CHAB has disappeared and cyanotoxin concentrations are below WHO guidelines. As a rule, a timely follow-up investigation is scheduled to ensure safety. The following comparison of the current and the Cyano+ workflow illustrates how fragmented information sources and largely manual assessment processes limit the potential for integrated risk assessment and forecasting. Addressing these limitations requires the comprehensive integration and structuring of relevant datasets, which is discussed in the following section.

Current Workflow:

Inspection → Sampling → Analysis → Report → Authority Decision

Cyano+ Workflow:

Monitoring Data → Integration → Risk Assessment → Forecasting → Decision Support

The advantages of the Cyano+ approach encompass an improved, shared situational awareness as well as a basis for faster decisions due to a common platform.

3.3. Integrated data

Building upon the identified limitations of the current workflow, particularly the fragmentation and lack of structured data integration, this section focuses on the systematic analysis and consolidation of relevant datasets. A comprehensive and harmonized data basis is a fundamental prerequisite for enabling reliable and timely data-driven decision support in the context of cyanobacterial risk assessment. Within the Cyano+ framework, multiple heterogeneous data sources are therefore identified, evaluated, and integrated to support monitoring, analysis, and prediction of CHAB events.

The data landscape relevant for cyanobacteria monitoring in Austria is characterized by a high degree of heterogeneity. Existing datasets are distributed across different systems, including laboratory information management systems, reports, tabular datasets and geospatial data formats. As highlighted in Section 3.2, these data are often only partially structured and lack consistent linkage between sampling sites, measurements, assessment, and management actions.

Within Cyano+, the relevant datasets can be grouped into several categories. In-situ observations, primarily provided by AGES, include analytical results on cyanobacteria, cyanotoxins, and supporting physicochemical parameters. These datasets provide high-quality reference information but are typically limited in temporal frequency and spatial coverage. Geospatial baseline data, such as digital elevation models, catchment boundaries, shoreline characteristics, land cover, and land use information, are sourced from BEV and enable spatial contextualization of monitoring sites.

Meteorological data, including air temperature, precipitation, wind speed and direction, solar radiation, and humidity, are provided by UBIMET as well as through numerical simulations (e.g., WRF, COSMO) generated within AIT projects. These datasets are essential drivers for modelling cyanobacterial dynamics. In addition, historical datasets, including CHAB occurrence records, bathing water reports, previous monitoring campaigns, and meteorological archives are used to support model development, calibration, and long-term analysis. A central challenge lies in harmonizing these heterogeneous datasets with respect to format, temporal resolution, and spatial scale. To address this, Cyano+ implements a structured data integration pipeline including data cleaning, normalization, temporal alignment, and semantic mapping into a unified data model. This integrated data foundation enables subsequent analytical processing and forms the basis for the system architecture described in the Section 3.4. For example, satellite-derived CHAB indicators can be combined with local meteorological conditions and historical monitoring records to identify environmental patterns associated with CHAB formation and persistence.

Table 1. Overview of integrated data sources in Cyano+

Data Category	Examples	Source	Temporal Resolution	Spatial Resolution	Role in System
In-situ & Observations	Cyanobacteria, cyanotoxins, water quality parameters	AGES	Low–medium	Point-based	Ground truth, validation
Geospatial Data	DEM, lake characteristics, land cover, land use	BEV	(Semi-) Static	High	Spatial context
Meteorological Data	Temperature, precipitation, wind	UBIMET, AIT	High (hourly)	Regional/grid	Model drivers
Climate Simulations	WRF, COSMO scenarios	AIT	Medium-High	Regional	Scenario analysis
Historical Data	Past bloom events, reports	Various	Low-medium	Aggregated	Model training

Source: authors

3.4. Concept of the overall system

Building on the data integration concept described in Section 3.3, the Cyano+ system is designed as an integrated decision support system (DSS) that transforms heterogeneous environmental data into actionable information for the monitoring and management of CHAB events. Figure 1 illustrates the core components of the Cyano+ system, including data sources, internal processing and analytical components, as well as interfaces to external systems.

A distinguishing feature of the Cyano+ DSS is the integration of environmental datasets that are traditionally managed separately. These include in-situ monitoring data (e.g., sensor measurements of water quality parameters, cyanobacteria and cyanotoxin observations), remote sensing data (e.g., satellite-derived indicators such as chlorophyll and CHAB-related indices), meteorological forecasts, hydrodynamic and ecological model outputs, historical CHAB records, and climate projections. By consolidating these heterogeneous datasets within a single analytical framework, the system enables cross-domain analysis and supports both short-term operational decision-making, such as early warning and risk communication, and long-term risk assessment and planning.

The system follows a modular architecture in which data acquisition, processing, analysis, and dissemination are logically separated but tightly interconnected.

At the data ingestion level, all data sources are integrated through standardized interfaces and directed into the DSS, ensuring a centralized integration point without direct dependencies on external systems. Given the diversity of formats (e.g., Excel, GIS, netCDF), preprocessing plays a critical role and includes validation, cleaning, temporal alignment, and spatial harmonization. This ensures data consistency and quality before further analysis. The system is designed to handle varying data resolutions and update frequencies, enabling both near real-time assessments and retrospective analyses. Based on this, a modelling level is implemented within the DSS, where analytical modules apply a combination of rule-based approaches, statistical methods, and machine learning techniques to detect patterns, assess risks, and generate forecasts of CHAB events. These modules are extensible, allowing new models or algorithms to be incorporated as scientific understanding and data availability evolve.

The results of these analyses are made available through operational system components that support user interaction and decision-making. A key component is the Emergency Map Tool (EMT), which provides spatially explicit visualization of monitoring data, model outputs, and risk indicators, enabling users to assess current conditions and identify potential hotspots. In addition, a Forum component supports communication and coordination among stakeholders, including environmental agencies, authorities, and system operators, facilitating information exchange and collaborative decision-making processes. The integration of sensor platforms (e.g., Badger Meter) further enhances the system by providing continuous, high-frequency monitoring data that complements traditional sampling approaches.

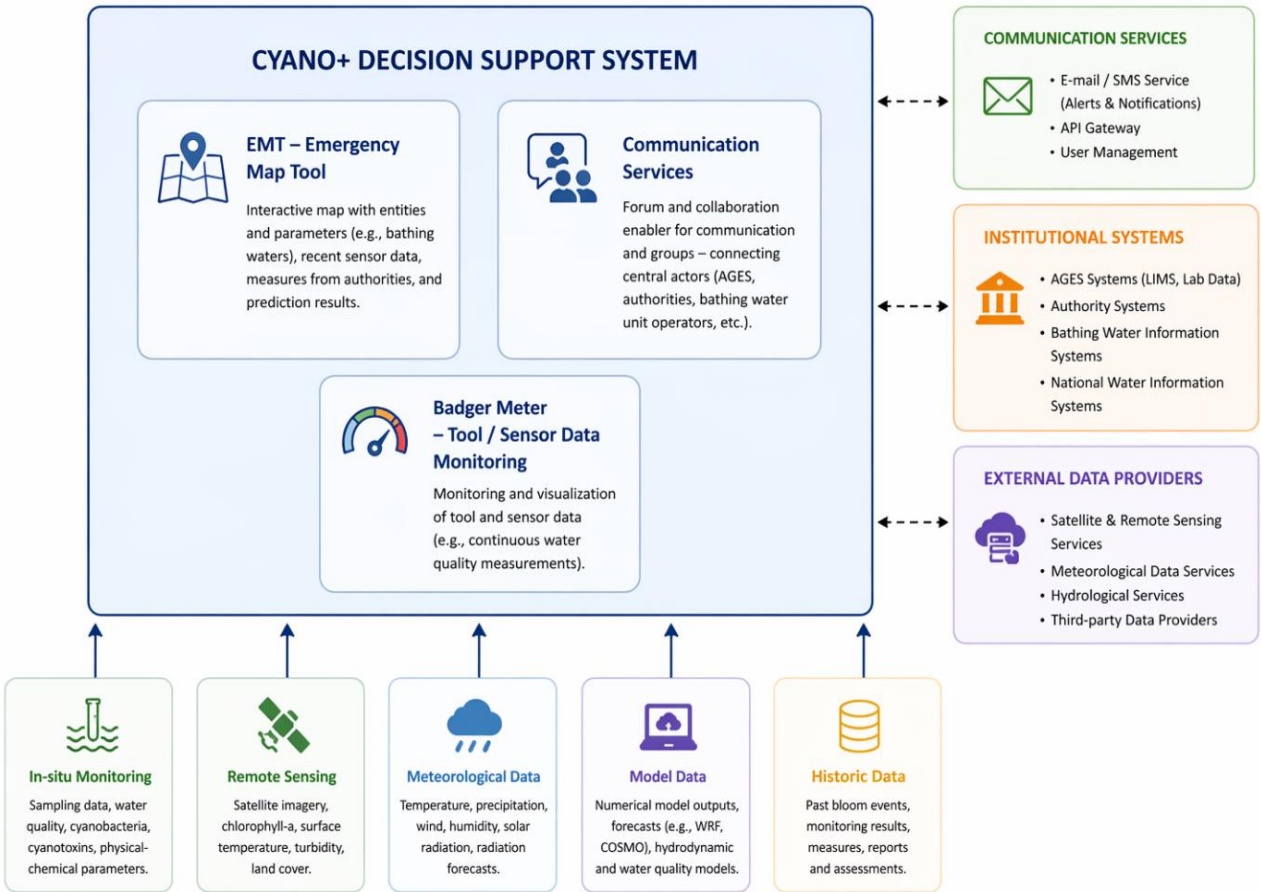


Figure 1. Component-based architecture of the Cyano+ system Source: authors

A decision-support layer emerges from the interaction of these components, providing functionalities such as risk indicators (e.g., CHAB probability levels), early warning mechanisms, and visual representations of spatial and temporal developments. The system supports the separation between technical assessment and formal decision-making, as described in Section 3.2, by providing structured and transparent information to decision-makers.

The processed information is also made available to external systems and stakeholders through clearly defined interfaces. Communication services enable alerting and notification functionalities, while institutional systems and external data providers can exchange information with the DSS and integrate results into their workflows. Importantly, all primary data flows are directed into the Cyano+ DSS, ensuring that it remains the central point of integration and intelligence.

A key design principle of the Cyano+ system is scalability and interoperability. The modular architecture enables the integration of additional data sources, the extension with new analytical models, and the transfer of the system to other regions. Standardized interfaces ensure compatibility with existing monitoring infrastructures and support long-term system evolution.

4. Discussion and conclusion

Climate change is expected to exacerbate the frequency, intensity, and duration of CHAB events. Elevated water temperatures increased nutrient runoff from agriculture and urban areas, and more frequent extreme weather events such as heavy rainfall followed by drought all create favourable conditions for cyanobacteria. In Austria, this could result in more frequent and widespread CHAB events in the coming decades, which could potentially increase the risk to public health, drinking water supplies, and local ecosystems. Currently, both cyanobacteria and cyanotoxins are examined as part of bathing water monitoring in Austria, but only on an ad hoc basis. However, there is currently no risk assessment for bathing water, groundwater, or subsequently for drinking water treatment plants related to the risks arising from cyanotoxins. Solutions for mitigating CHAB events using physical, chemical, and biological approaches are currently being sought. Cyano+ addresses these contaminants from a hazard perspective. The project develops methods for risk assessment and evaluation of “emerging contaminants/pollutants” such as cyanotoxins (cyanobacteria) in bathing waters, as well as a forecasting method for CHAB occurrence and a fluorescence-based sensor for detecting cyanobacteria. A core element enabling data fusion and integration is the Cyano+ Decision Support System (DSS). The Cyano+ DSS converts analytical findings into targeted, actionable insights for stakeholders such as environmental agencies and authorities. It provides risk indicators (e.g., CHAB probability), early warnings, and clear visualizations of spatial and temporal trends.

The Cyano+ project makes an important contribution to the reduction of public exposure to cyanobacteria and cyanotoxins by providing recommendations for monitoring measures and developing a corresponding risk assessment for various systems. It enhances the resilience and sustainability of water resources by enabling early action through the methods developed. These methods help maintain the ecological status of water bodies in line with the Water Framework Directive, reduce negative economic impacts, and facilitate the implementation of appropriate response measures. The Cyano+ DSS is designed modularly, allowing for system adaptations or extensions to address additional use cases, such as qualitative or quantitative threats to drinking water supply.

Acknowledgement

The research leading to these results has received funding from KIRAS Cooperative R&D Projects 2024 Cyano+ with project number FO999914131_02082024_144926951. Copilot from Microsoft, GPT from OpenAI, Gemini from Google and LeChat from Mistral were used to retrieve information on existing Decision Support Systems and to summarise and optimise parts of this paper.

References

Anderson, M., Kaoru, Y., & White, A.W. (2000). Estimated Annual Economic Impacts from Harmful Algal Blooms (HABs) in the United States - Technical Report WHOI 2000-11 (97 p.). Woods Hole Oceanographic Institute, Woods Hole, Mass.

BGBl. I. 249 (1976). Bundesgesetz über Hygiene in Bädern, Warmsprudelwannen (Whirlwannen), Saunaanlagen, Warmluft und Dampfbädern sowie Kleinbadeteichen und über die Wasserqualität von Badegewässern (Bäderhygienegesetz – BHygG. https://www.ris.bka.gv.at/GeltendeFassung.wxe?Abfrage=Bundesnormen&Gesetzesnummer=10010382

BGBl. II. 349 (2009). BGBl. II Nr. 349/2009 Verordnung des Bundesministers für Gesundheit über die Qualität der Badegewässer und deren Bewirtschaftung (Badegewässerverordnung – BGewV).

Codd, G.A., Young, F.M., & Utkilen, H.C. (2005). Europe: Cyanobateria, cyanotoxins, their health significance and risk management. In Codd, G.A., Azevedo, S.M.F.O., Bagchi, S.N., Burch, M.D., Carmichael, W.W., Harding, W.R., Kaya, K., & Utkilen, H.C. (eds.) CYANONET. A Global Network for Cyanobacterial Bloom and Toxin Risk Management Initial Situation Assessment and Recommendations.HP-VI, Technical Documents in Hydro, 76, 7193. UNESCO, Paris.

Digital Water.City – Machine-Learning Based Early Warning System for Bathing Water Quality. https://www.digital-water.city/solution/machine-learning-based-early-warning-system-for-bathing-water-quality/

DIRECTIVE 2006/7/EC. (2006). Directive of the European Parliament and of the Council of 15 February 2006 concerning the management of bathing water quality and repealing Directive 76/160/EEC, https://www.parlament.gv.at/gegenstand/XXVII/EU/73595

Istvánovics, V., Honti, M., Torma, P., & Kousal, J. (2022). Record-setting algal bloom in polymictic Lake Balaton (Hungary): A synergistic impact of climate change and (mis) management. Freshwater Biology, 67(6), 1091–1106.

iSWIM. Bathing Water Quality Monitoring in the Black Sea, Copernicus Marine Service. https://marine.copernicus.eu/services/use-cases/iswim-bathing-water-quality-monitoring-black-sea

Raeder, U. (2022). Bayrische Seen im Klimawandel.-Bayerisches Staatsministerium für Umwelt und Verbraucherschutz (108 p.). München. Sene, K. (2025). Hydrometeorology – Forecasting and Applications, Springer Cham. https://doi.org/10.1007/978-3-031-58269-1

Shutler, J.D., Warren, M.A., Miller, P.I., Barciela., R., Mahdon, R., Land, P.E., Edwards, K., Wither, A., Jonas, P., Murdoch, N., Roast, S.D., Clements, O., & Kurekin, A. (2025). Operational monitoring and forecasting of bathing water quality through exploiting satellite earth observation and models: The AlgaRisk demonstration service, University of Exeter. https://ore.exeter.ac.uk/articles/journal_contribution/Operational_monitoring_and_forecasting_of_bathing_water_quality_through_exploiting_satellite_earth_observation_and_models_The_AlgaRisk_demonstration_service/29708111?file=56957477

SWIM: AI – Early Warning System for Pollution at River Bath Sites – Digital Water.City, Kompetenzzentrum Wasser Berlin GmbH. https://kompetenz-wasser.de/media/pages/forschung/dienstleistungen/fruehwarnsystem-fuer-badegewaesserqualitaet-datengetriebene-vorhersage-zur-gewaesserqualitaet-von-badestellen/a8c92dc213-1658847657/swim-ai.pdf

Valeriani, F., Carraturo, F., Lofrano, G., Volpini, V., Izzo, M.G., Bruno, A., Guida, M., & Romano Spica, V. (2024). Algae in Recreational Waters: An Overview within a One Health Perspective.Water, 16, 946. https://doi.org/10.3390/w16070946

CYBERRESILIENCE FOR SECURITY

CRYPTOGRAPHY IN A POST-QUANTUM WORLD

Michael Sonntag

Johannes Kepler University Linz
michael.sonntag@ins.jku.at

DOI: 10.35011/IDIMT-2026-43

Keywords

Cryptography; post-quantum; application areas.

Abstract

This paper surveys the practical transition from conventional to post-quantum/quantum-safe cryptography (PQC) across contemporary infrastructures. Large-scale quantum computers impact cryptographic primitives differently: conventional asymmetric encryption/signatures will become insecure, while symmetric ciphers largely remain viable and hash functions face only modest weakening. Therefore, adaptations to existing as well as future systems are needed. Depending on the application area, "harvest-now, decrypt-later" may either be a real threat or not, so urgent action is required only in those areas, where once security has been broken it cannot be reconstructed, while others may have more time to adapt (e.g. where data has a very short lifetime anyway). The paper discusses a long list of potentially affected systems, and what special challenges implementing post-quantum cryptography poses for them, and where upgrading needs to be prioritized. A problem identified as remaining unsolved is "grandfathering": what to do with devices that simply cannot be updated, e.g. because the cryptography is integrated into their hardware.

1. Introduction

Encryption has served humanity for a long time. Almost all of that time it was (and remains) symmetric cryptography, where encryption and decryption employ the same key. Comparatively recently asymmetric cryptography was developed, where these operations use different but mathematically related keys. However, scientific advance does not stop and quantum computing came along. This might (and probably will rather sooner than later!) render some variants of cryptography useless in the sense, that it will be easy to break. This is specifically a problem for asymmetric cryptography – which despite its young age is indispensable for today's infrastructure: symmetric cryptography works between a small (typically two members only) group of persons, trusting each other with a shared secret. Generally, this secret can be any possible value, so "improved mathematics" is not an (as seen in practice: fundamental and real, but here rather limited) problem. Asymmetric cryptography however depends on the two keys having a special mathematical relationship – and unfortunately the algorithms currently used will no longer be secure when real quantum computing has been implemented.

While simulations of quantum computers do work, they are not useful for real-world applications. But actual physical quantum computing devices have been built (in a tiny size), and have been continuously enlarged to small sizes in the last years. Unfortunately (from the point of cryptography users) it does not seem that there exists a fundamental problem to further enlarge these, so they must

be expected to be able to work on actual real-world problems in the future. This does not mean success is guaranteed (there might still be as of yet unknown problems/insurmountable obstacles ahead), but it looks very likely and estimates for a timeline exist (albeit with uncertainties and only as probabilities; see e.g. BSI TR-02102-1: RSA, ECDH recommended until 2031, or Mosca/Piani 2025: breaking RSA-2048 within 24 hours - 6 of 25 experts say 95% probability within 15 years).

Consequently, it is necessary to prepare for this likely and not too-far-off future. The reason for this is twofold: imagine e.g. an electronic signature that is intended to prove a land transfer (signature on a contract of sale). While the contract was concluded in the past, its validity may be questioned for a long time (Austria: statute of limitations is typically 30 years; in some cases even 40 years). Therefore, it being impossible to falsify today, but definitely an option in e.g. 15 years is still a problem: the amount of money or the parcel of land could be changed - or it could be claimed to be invented as a whole, as a valid signature can be produced for an alternative version. And it is impossible to prove on its own which of the two variants is the original/binding one (or that not both are falsified). As a consequence, evidence valid at the current time gets suddenly, and especially irrevocably, useless later. The second reason is called "store now – decrypt later": while breaking encryption is impossible now, at a later point in time this might change - and storing data is cheap. Its content might still provide useful information (or evidence/blackmail material/…) later on. So once an algorithm has been broken (with wider availability of the required soft-/hardware), these cryptographic schemes will be useless (retroactively and for the future) – and as this pertains to all current asymmetric cryptography, this would cause huge problems.

It is therefore necessary to investigate now, which cryptographic schemes will likely become obsolete to develop (to a large degree already completed) alternatives, and implement these (currently sorely lacking) in preparation for the date when they are required to *have been implemented before*. This paper contributes to the second part: a systematic investigation which elements in what types of applications need to be adapted.

2. Which cryptography is affected how?

Several different kinds/types of cryptography exist, which are discussed below with their potential post-quantum problems (if any; see also Joppich 2025):

- Asymmetric cryptography: RSA and ECC are affected most, as these must be considered completely broken if a large enough quantum computer is available (through using Shor's algorithm). This affects the algorithm itself and so applies to both encryption and signatures.
- Symmetric cryptography: For these there seems to be no general problem. However, key length might have to be doubled (see Westerban 2025 for why this could be optional, as quantum attacks might be impractical) to retain the same strength (reduced by Grover's algorithm). As a consequence, AES is only slightly affected and AES with a key length of 256 bits should still be sufficient. However, it might be useful to start research into/standardization for variants with longer keys.
- Key agreements algorithms: These are typically (Diffie-Hellman as well as its Elliptic Curve variant – DH and ECDH) based on similar mathematical problems as asymmetric cryptography and therefore must also be considered imperilled/broken in the future. They are rarely used alone but typically combined with signatures to prevent Man-in-the-middle attacks and therefore "twice" as problematic.
- Hash algorithms: Quantum technology does not seem to (Bernstein 2009) impact them significantly; they are slightly weakened similar (but it seems not identically to) symmetric

cryptography and therefore longer variants should be preferred. E.g. the recommendation (CNSA 2.0; NSA 2024) is to use at least 384 bits (replacing 256 and 384 bit hash functions; SHA3-256/-384 → SHA3-384) or 512 bits ("replacing" 512 bits: SHA3-512 → SHA3-512). Here the same applies as for AES: longer variants should perhaps be developed to ensure a sufficient security margin for the future.

- Key derivation functions: Deriving a key from a password must be adapted (but not replaced), as different kinds of keys are required for post-quantum cryptography. However, care is necessary in this category, as variants of "combining keys" exists and may have different requirements: e.g. hybrid approaches can combine classical and post-quantum keys into a single symmetric key. Using hash algorithms to combine/create keys is similar as discussed above: only slightly affected.

However, while some new algorithms exist for each affected cryptography systems, this is not the only problem when moving on. The following issues will also have to be taken into account:

- Protocol compatibility: If a new system is developed from scratch, new algorithms can trivially be integrated. But existing systems need to be upgraded too. For them additionally protocols employed (because of old/not-yet-upgraded counterparts) must support selectable cryptography: conventional and post-quantum simultaneously. Some systems (e.g. IPSec) already support algorithm agreements and therefore the addition and use of new ones, but not every system can do this. As a consequence, initiators will have to try the "new" version first (and listeners respond to "old" versions too) – which opens up the possibility for downgrade attacks. But these are difficult to protect against (e.g. using - and maintaining! - lists of known-to-be-upgraded counterparts for which only PQC is accepted).
- Legacy systems: This ties in with the previous issue - systems that cannot be upgraded, e.g. because the vendor company no longer exists or no longer supports that system, or where the libraries used cannot be upgraded without also changing a lot of other code etc. There does not seem to be a good solution for this problem, but e.g. tiny "wrapper-boxes" creating a secure tunnel in front of them could be used. This is easy for closed groups and pre-determined static communication links, but would require serious international standardization for open systems (compare to IPSec) or turning of all encryption at the source so these middleboxes can inspect the traffic to extract the recipient address(es). Applicability is therefore probably limited to comparatively small groups of elements with a known set of counterparts. Note that newer systems could natively implement this, but still face the problem of downgrade attacks (other than the "fixed line with PQ-encryption" approach with middleboxes). It is also important that mutual authentication for deciding whether post-quantum security is required does not help here: it would have to take place via non-post-quantum means (i.e. conventional RSA/ECC certificates) as these are the only algorithms old systems can provide, and these will not be secure at that time anymore either. A potential (but limited, problematic etc) alternative could be random nonces: manually (or securely via modern systems) distributed and then used for one-time authentication.
- Larger key sizes: While some of the unaffected algorithms might have to increase their key length, at least some post-quantum algorithms require massively longer keys, sometimes several kilobytes (ML-DSA: 1-3 kB) or even longer. As such this is not necessarily a problem (memory is usually plenty; but can be problematic on small devices with limited memory, e.g. IoT), but databases might have to be changed to accommodate these new lengths, as well as variables for handling them etc. Additionally, today most network connectivity is using the IP protocol, which has a maximum packet size of 1500 bytes: while RSA/EC keys easily fit inside a single packet, especially when containing some other protocol data too, a post-

quantum key might no longer fit. Consequently, it must be split over multiple packets, reducing performance and causing perhaps problems on end devices if some fragments are missing. Attacks replacing parts of the key now also become possible (e.g. through replay) and new/additional integrity measures will be required to securely tie all key fragments together.

- Large signature sizes: Similar to long keys, signatures also increase in length. As they are typically protocol content (as opposed to initialization/setup) data this is of lesser consequence and usually not a problem (but see certificates below). Similar as above, storage needs to be considered (especially for fixed-length messages) as well as potentially bandwidth (for frequent individually signed messages: consider a chat application). Length ranges from 666 bytes (Falcon-512; Prest et al. 2020) to almost 50 kB (SPHINCS+-SHA2-256s; Aumasson et al 2022). However, if certificate chains are longer, multiple such signatures might have to be exchanged (e.g. assuming a chain of length 3 → at least 2…120 kB), significantly increasing transmission time and in sum even storage space.
- Longer computation/hardware support: While some algorithms are very fast (even faster than RSA/ECC), some are also significantly slower for either signing or verification, or for both operations. This is not only a problem as such, but also regarding hardware support. E.g. CPUs today often include specific instructions to implement parts of the crypto-operations in hardware – which is at least at the moment not true for these new algorithms and will never change for legacy hardware. This is more an implementation problem for cryptographic libraries, but still must be considered as a weakness, as a more complex implementation increases doubts about its security. Hardware support is also difficult to implement and therefore potentially delayed significantly even after standardization, as it is complex to avoid side-channel attacks (not all of which might be known; see below), especially when using general-purpose instructions.
- Lifetime: Some data must be secured for a long time (see introduction), but similar to asymmetric cryptography this is a trapdoor. Security must have been present during the whole previous time, for any even brief lapse might have led to a falsification during that time window. I.e., once security has expired, it cannot ever be reconstructed. In some circumstances this might be solved through extra-cryptographic means, e.g. re-digitizing physical contracts and signing the scans with a new algorithm, but from a digital point of view this is simply "new" data – so this approach will not work for *purely* digital content, e.g. contracts without a paper variant concluded solely via digital signatures. Here some side aspects are relevant: multiple signatures can be generated easily (in case one algorithm is assessed as insecure later on), data can be signed aggregated (not signing every single contract separately but a whole archive at once), and they need not be signed by the original person/s (like for a contract by one/both original signers, but instead by some "guarantor", e.g. a public notary). But they do require a secure timestamp, as this ensures the PQ-secure re-signing took place before the invalidation time of the conventional algorithm signatures. Whether this timestamp is an actual cryptographic timestamp or a trail of existence (e.g. a hash value of the signature printed on paper and physically dated and signed by several persons) doesn't matter.
- Random number generator: For classical asymmetric cryptography random numbers are only required for generating keys. However, some post-quantum algorithms require random numbers for signature generation too. This can be a problem for (some/old) virtualized systems or small devices, where good random numbers might be unavailable. Virtualization today typically passes physical random number generation features of CPUs through to containers/VMs, but IoT devices might not have any reliable random source at all. This must

be considered when selecting algorithms, as e.g. certificates might be generated on "good" systems, but then distributed and used on "weak" ones (not a problem for RSA/ECC).

- Less experience/research: Whether the new algorithms are really secure or not, is not as clear as for those, where scientists performed decades of research. The same applies to a lesser degree to their implementation: for conventional algorithms lots of side-channel attacks and prerequisites are known and how to avoid them. Whether the new algorithms open up new kinds of attacks is unknown (see e.g. Qiu/Aysu 2025 for an example of an implementation problem of a known type of sidechannel: power measurement during key generation allows a 99.99994654% key recovery success rate for FALCON-512).

To sum up, lots of new problems occur and predictability is low. As a consequence, "crypto agility" will be especially important in the near future. "Crypto agility" is defined as the ability to switch algorithms easily and quickly, so replacements (both from the algorithm and the implementation point of view) are simple and quick to perform without side-effects. This implies e.g. that protocols support selection/negotiation of the algorithm to be used, and few assumptions on elements (e.g. keys and signatures should be stored as BLOBs without any size assumptions, so that their actual length/content/encoding is irrelevant).

3. Types of affected systems

Which kinds of systems or devices will be affected by quantum cryptography problems and therefore need to be adapted/prepared? Some of them are widely known and cannot really be influenced by most companies (e.g. hardware support in CPUs: signatures built into them to prove authenticity, like SGX), but the implementations within e.g. OT devices, VPNs used for home office, or TLS connections are important for many companies and individuals. Relevant and affected by post-quantum problems at least to some degree are the following systems:

- Secure boot systems: Signatures of OS boot sectors/files in the BIOS are very important, because if they can be falsified to still match, attackers can replace them with tampered versions and verification will still succeed (or produce a "valid" update for the BIOS with new signatures). Here a longer lead-in time is required, as while business computers are exchanged every few years, many companies also use quite old systems. This area is especially problematic, as attackers need only forge a single malware-infected/manipulated bootloader, to be able to successfully attack potentially millions of devices. While updates seem reasonable for servers, end-user devices rarely receive (and users install) bios updates - probably none at all except if distributed via OS updates (Windows, Apple). Technically challenging is, that this signature must be verified through the BIOS/TPM, which not only needs firmware updates (both verification code as well as certificate data: BIOS), but might require hardware adaptations too (TPM; see below). However, as long as updates are available and installed early enough, even legacy systems can be upgraded.
- CPUs (e.g. Intel SGX): CPUs - at least of the server variety - often include cryptographic keys to support secure computation in various forms. Often these are tied to certificates proving that the device is an actual physical CPU (the randomly created private key is burnt into the hardware on manufacturing and the accompanying public key is signed by the manufacturer) and not a software emulation - which might behave differently. Updates are here impossible as key and algorithm are unchangeable; only replacing the whole CPU for a compatible newer one works. As such features are typically not present in end-user devices or even workstations, this poses a smaller problem: servers are exchanged regularly and even then these features

seem to be used comparatively rarely. Still, the lead time is significantly, as not only the storage but also the handling (=cryptography) needs to be implemented in hardware.

- Trusted Platform Modules (TPMs): These need to be significantly expanded, as they must now support additional algorithms and enlarged storage. Some TPMs are plug-ins (e.g. pin header), i.e. they could be replaced, but most are directly built into appliances, preventing their replacement. Consequently, this renders these devices obsolete and therefore needs to be addressed soon so that capable products will be prevalent when the need for switch-over occurs (especially companies might use hardware for a long time, so at least 5 years in advance will be necessary). Also, old devices will become obsolete, as no upgrade path exists. As TPMs may also be used for DRM (Digital Rights Management), media players will be affected too (very problematic for copyright owners, as every "old" material – stored with conventional cryptography DRM - can then be liberated from its protection).
- Hardware tokens (e.g. for bank transfer authorizations or MFA login): Depending on their inner working (signature or some variant of TOTP) they might require replacement, as "updating" them, like all hardware, is unrealistic. This especially applies e.g. to Fido tokens. Unlike PCs or CPUs, distributing new tokens is easily feasible, even on a short notice - as long as they are physically available (small, cheap, easy to manufacture). Regarding TOTP: while improvements have been proposed (Kurariya/Bodhankar/Shah 2025; especially regarding distributing the shared secret, which is often secured by a – currently conventional – signature), TOTP itself is based on hash values and therefore only affected to a very limited degree.
- Unlock tokens, e.g. car keys, garage door openers, alarm system remote controls (not a pure security device, but requiring at least some protection): If they are based not on a shared secret but asymmetric cryptography, these are affected and difficult to replace. This is based less on the key (see hardware tokens before), but rather the support within the car/garage/etc. Note that here typically wireless communication is used, which means that any increase in signature length can be very problematic (duration of unlock, changes in required speed, potential for relay attacks). This applies to both simple "unlock" keys, as well as immobilization systems blocking activation (e.g. ignition for cars) if the key is not close by (see e.g. Embitel 2025 – very recent but devoid of any reference to post quantum problems!). Because of their potentially long lifetime (and sometimes even increased value over time - Oldtimers) this can be a significant and urgent problem.
- Passkeys: Passkeys as replacement for passwords can be adapted rather easily, as just a different type of key needs to be used. Also, they are never entered/handled/… directly by humans. Backwards-compatibility may be an issue, as well as additional keys, e.g. for recovery, which are not yet post-quantum secure. Protocols already allow algorithm specification, so extending those should be straightforward. Hardware tokens for storing them will however require new devices (see above). One problem is, that existing keys are not easily "updated": before they become invalid, every single site they provide access to must be visited (!), a new key individually enrolled there, and the conventional old key removed: no automatic/offline upgrade possible as interaction is mandatory.
- Password/key vaults: These already need to store various cryptographic keys (and usually synchronize them), so new types of keys with different length are trivial to integrate. Protection of the storage itself is encryption, so should be affected to a smaller degree. Only synchronization between devices/servers, where identifying the other end is essential, requires adaptation. For the contained keys see Passkeys; passwords are not affected.

- Timestamps: The timestamp itself is unaffected, but its signature (the trusted third party confirming that the data, respectively its hash, was presented to her at that time) must be replaced with a post-quantum version. This requires a changed issuing process (typically using hardware security modules) and new verification libraries, but no inherent difficulty exists. As timestamps may occur in various and numerous contexts, the in some algorithms larger signature length may be a problem in some cases. Note that timestamps often need to remain valid for a long time, so implementation is more urgent and methods for extending the validity of old timestamps may have to be used. A trivial extension can be implemented through creating a new PQ-timestamp over the old timestamp at a date before the old one becomes insecure: it provably existed at a time when it could not yet have been falsified.
- Web browsers: Their update cycle is very short and rolling out selected post-quantum cryptography has already started (Chrome/Edge: Version 131; Firefox: Version 132; Safari: Version 26; see Cloudflare 2026). Problematic are especially certificates (see below). Through the automatic update mechanism, this area should theoretically be a tiny problem. However, web browsers are often integrated into other devices too, e.g. Smart-TVs, mobile phones, tablets etc. For some of these devices updates are rarely installed - or even provided by the manufacturer at all. It must therefore be expected that for a significant length of time devices will be in use that do not support adequate protection mechanisms. These will either have to be excluded (i.e. rendered inoperative regarding web browsing), or a conventional fallback must be supported by websites for them. For some devices (e.g. online movie shops for Smart-TVs or game consoles) this might be feasible (insecure, but at least no payment information must be transmitted – e.g. only orders to buy/rent something for an account with the payment information associated therein; login is however problematic as transmitted insecurely; custom implementations of e.g. zero-knowledge proofs via JavaScript could overcome this, but these would have to be again installed locally to avoid code manipulation during download), but premature obsolescence must be expected. Here again an adequate preparation time is needed, but especially for older game consoles this might also render them (not only the console as such, but e.g. online games too) effectively unusable: no updates are available, but an insecure online connection would be mandatory for use.
- Certificates: These are used in a large number of different contexts to prove the identity of an entity, often solely some device like a server. A problem with certificates is, that they often must be presented in advance of content communication, i.e. before the needs/capabilities of the clients are known. Consequently, for some time hybrid certificates containing both conventional and post-quantum keys will be necessary to avoid complex protocol changes. Combined with the longer signatures/public keys of post-quantum algorithms this will lead to significantly larger certificates. The enlarged size can be a problem for smaller devices, as these might have a limited disk/persistent storage size and therefore can now only store a smaller number of certificates (or not at all in case of very limited space for a single very large certificate). Old and incorrect implementations may even decline a session start as invalid (Schaumann 2025) when using hybrid certificates. Additionally, when using post-quantum or hybrid certificates with TLS to identify a webserver this leads to longer connection setup times, as more packets need to be transferred before a secure connection can be established (obviously dependent on the connection speed; at least for some post-quantum algorithms this seems to be a small problem: Jarvis, 2022). Recent advances (e.g. TLS 1.3) reduce this problem by performing identification only once and then using session resumption. Note the privacy implications of that, as clients need to remember (=store) data about servers they recently connected to, as well as servers then recognizing clients as the same entity. Google experiments with different kinds of "certificates" (Merkle Tree Certificates) to avoid such size

issues, especially for longer certificate chains (Chrome Secure Web and Networking Team 2026). Another issue with certificates is, that they not only contain keys, but are also signed themselves: CAs must support creation of these signatures and here too hybrid (= at least two) signatures will be necessary. And while a website might provide a post-quantum certificate, anyone can later create a valid "classical-only" certificate by breaking the private key and then intercepting requests and stating "the website cannot perform any post-quantum cryptography", i.e. a downgrade attack. Therefore, additional mechanisms to declare "we support PQ" are required (Benjamin/DeBlasio 2026). The certificate ecosystem also encompasses certificate transparency logs, which must be adapted too (very few operators; transition/compatibility will probably be the biggest challenge there).

- Virtual Private Networks (VPN): These typically require some authentication of the other side, often via private keys or certificates. In the latter case these systems are affected (those using TLS like OpenVPN are similar to web surfing and obviously affected, but also secured alike). Many systems depend on symmetric keys and/or pre-shared password and so are in this aspect secure. Probably all of them however will use some key agreement for dynamically changing keys and therefore must be upgraded. For some (e.g. IPSec) this should be comparatively easy, as they already have a method for algorithm selection built in, or depend on other protocols (OpenVPN: TLS). Wireguard however is PQ-insecure, but through adding a pre-shared key (which can be agreed upon by both sides via a quantum-secure algorithm) it can be made secure (Hashimoto et al 2026) without too many changes.
- Payment cards/Smartcard: Such cards are on the one hand problematic, as they contain a very weak processor that is additionally energy-constrained (no own power source) and quick results are required, on the other hand they are regularly exchanged. So "rollover" should be comparatively easy and cheap (but may still need several years, e.g. for credit cards and especially SIM cards), once the required algorithms are implemented in hardware. Because of extensive and complex validation, this will take some time too. However, post-quantum algorithms as such may be problematic for them because of the resource constraints. This might render some post quantum algorithms which are no problem for a PC unsuitable for these scenarios.
- OT devices: Devices used in industrial control systems possess an especially long lifetime, typically 20 years or longer. Consequently, according to some estimates, they have already "lost", as many already deployed non-post-quantum-ready devices will still be in normal and intended use after their signatures can be broken (assuming sufficiently large quantum computers arrive at the earliest estimates). This might be mitigated by being physically hard to reach (a questionable assumption in many industries) and completely separated from the Internet (today unrealistic, and coupled with internal attacks only of limited use too). For them also updates might be difficult: limited (and potentially used to that limit) computation power as well as storage might require physical replacement. Also, updates are complex and rare (the industrial system might have to be turned off/put into maintenance mode, advance testing is difficult etc). Note that all devices that are communicating with each other must be pairwise upgraded simultaneously (or support fallback methods) – with several manufacturer's devices being used in a site as additional challenge. For these products therefore the capability for later upgrades via software is of utmost importance: they cannot be equipped with standardized (important, as such system may have to be officially inspected/verified/certified/…) post-quantum cryptography now, but they must be upgraded later in their lifetime. Fortunately, such update/upgrade functionality is required by the Cyber Resilience Act of the EU too, so such features must be integrated very soon (full implementation: 11.12.2027) anyway.

- Cryptocurrencies: For them the “harvest now, decrypt later” attack is a massive problem, as some wallets remain unchanged for many years (Ethereum 2025). Compromising them would get attackers access to significant amounts of money; also their public keys are often exposed to the world by design. Additionally, a successfully forged history would cast doubt on the validity of a chain as a whole. But many cryptocurrencies are themselves based on (only slightly affected) hash values, while the (typically elliptic curve) signatures (or encryption) are used only in different aspects (e.g. signed proposals, consensus voting, transaction signatures, and especially wallets).
- Also relevant, but not discussed here: mobile phone systems (connection between devices and base stations), M2M communication of industrial systems (as far as not Internet- or mobile phone-based), DNSSec (small signature size is an important aspect), digital identities/wallets (e.g. ID Austria, mobile driving license etc; these also will need to be investigated and will require updates, as they are directly based on digital signatures) and probably even more.

Considering the investigation, the priority needs to be on long-living devices: the longer a system will be in use, the earlier post-quantum security must be introduced. Also prioritized must be hardware: software is typically quicker to produce and replace, than hardware, especially if it cannot be replaced separately from a larger device. As an intermediate measure to buy time, the option to change algorithms as well as update functionality must be introduced.

4. Summary and outlook

Progress of the transition from classical asymmetric cryptography towards post-quantum secure algorithms has been made, especially regarding the development and standardization, and first implementations (key exchange in Web-Browsers) are being rolled out. However, these cover only a small part of the landscape to upgrade and large areas remain as yet untended. As discussed above, software can (at least in widespread end-user areas) be updated comparatively easily – as soon as available and tested. But everything in a “device” and obviously all hardware itself needs a much longer preparation and earlier roll-out, as changes are more complex, need new hardware that must be designed, tested, manufactured, approved and integrated etc. And as devices have a (much; potentially even second-hand use) longer lifetime compared to most software, they require a matching longer transition period. As a company it is therefore important to start investigating where exactly changes will be needed, where vendor support or new devices are required, and then create a timeline for adaptation. This is especially important as the date when a sufficiently large quantum computer will be available cannot be determined precisely, but only estimated.

References

Aumasson, J.-P., Bernstein, D.J. Beullens, W., Dobraunig, C., Eichlseder, M., Fluhrer, S., Gazdag, S.-L., Hülsing, A., Kampanakis, P., Kölbl, S., Lange, T., Lauridsen, M.M., Mendel, F., Niederhagen, R., Rechberger, C., Rihneveld, J., Schwabe, P., & Westerbaan, B. (2022). SPHINCS$^+$. https://sphincs.org/data/sphincs+-r3.1-specification.pdf

Benjamin, D., & Deblasio, J. (2026). Post-Quantum HTTPS Authentication Roadmap. https://www.chromium.org/Home/chromium-security/post-quantum-auth-roadmap/

Bernstein, D. J. (2009). Cost analysis of hash collisions: Will quantum computers make SHARCS obsolete? https://cr.yp.to/hash/collisioncost-20090823.pdf

BSI TR-02102-1 (2026). Kryptographische Verfahren: Empfehlungen und Schlüssellängen. https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/Publikationen/TechnischeRichtlinien/TR02102/BSI-TR-02102.html

Chrome Secure Web and Networking Team (2026). Cultivating a robust and efficient quantum-safe HTTPS. https://security.googleblog.com/2026/02/cultivating-robust-and-efficient.html

Cloudflare (2026). PQC support. https://developers.cloudflare.com/ssl/post-quantum-cryptography/pqc-support/

Embitel (2025). Managing Cryptographic Keys in Automotive PKI: A Guide for OEMs and Security Architects. https://www.embitel.com/blog/embedded-blog/managing-cryptographic-keys-in-automotive-pki

Ethereum (2025). Future-proofing Ethereum and crypto quantum security. https://ethereum.org/roadmap/future-proofing/

Hashimoto, K., Katsumata, S., Niot, G., & Wiggers, T. (2026). Revisiting PQ WireGuard: A Comprehensive Security Analysis With a New Design Using Reinforced KEMs, IEEE Security & Privacy 2026. https://thomwiggers.nl/publications/pq-wireguard/

Jarvis B. (2022). How to tune TLS for hybrid post-quantum cryptography with Kyber. https://aws.amazon.com/blogs/security/how-to-tune-tls-for-hybrid-post-quantum-cryptography-with-kyber/

Joppich, W. (2025). Which Post Quantum Cryptography (PQC) Algorithm Should I Use? https://www.fortanix.com/blog/which-post-quantum-cryptography-pqc-algorithm-should-i-use

Kurariya, P., Bodhankar, J., & Shah, A. (2025). Enhancing Time-Based One-Time Password (TOTP) Authentication with Post-Quantum Cryptography Using CRYSTALS-Dilithium, 2025 IEEE International Conference on Blockchain and Distributed Systems Security (ICBDS), Kolhapur, India (pp. 1-7). doi: 10.1109/ICBDS67396.2025.11373926

Mosca, M., & Piani, M. (2025). Quantum Threat Timeline Report 2025. https://globalriskinstitute.org/publication/quantum-threat-timeline-report-2025b/

NSA. (2024). The Commercial National Security Algorithm Suite 2.0 and Quantum Computing FAQ (v2.1). https://media.defense.gov/2022/Sep/07/2003071836/-1/-1/0/CSI_CNSA_2.0_FAQ_.PDF

Prest, T., Fouque, P.-A., Hoffstein, J., Kirchner, P., Lyubashevsky, V., Pornin, T., Ricosset, T., Seiler, G., Whyte, W., & Zhang, Z. (2020). Falcon: Fast-Fourier Lattice-based Compact Signatures over NTRU. https://falcon-sign.info/falcon.pdf

Qiu, J., & Aysu, A. (2025). SHIFT SNARE: Uncovering Secret Keys in FALCON via Single-Trace Analysis, https://arxiv.org/html/2504.00320v1

Schaumann, J. (2025). Post-Quantum Cryptography Implementation Considerations in TLS. https://www.akamai.com/blog/security/post-quantum-cryptography-implementation-considerations-tls

Westerban, B. (2025). State of the post-quantum Internet in 2025. https://blog.cloudflare.com/pq-2025

MULTIPLE ETHICAL ASPECTS AND LIMITS OF ICT RESILIENCE

SELECTED ETHICAL ASPECTS AND LIMITS OF ICT RESILIENCE

Anton Lisnik

Slovak University of Technology in Bratislava
Institute of Management
anton.lisnik@stuba.sk

DOI: 10.35011/IDIMT-2026-55

Keywords

ICT resilience; ethics; cybersecurity; privacy protection.

Abstract

Today's times have encountered limits associated with the perception of privacy and its protection from an ethical point of view and the vulnerability, or robustness, of ICT systems. In addition to the technical and software problem, it is also an ethical problem, which has various overlaps changing the view of the privacy of ICT users, the discretion of information and, for example, the right to privacy. It is necessary to find the right level of mutually functionally respectful rules. The study examines the ethical aspects of ICT resilience based on empirical data from 50 respondents. The research focuses on the relationship between security and privacy. The study emphasizes the importance of ethical aspects in modern ICT systems.

1. Introduction

ICT systems form the backbone of modern society and support services in healthcare, finance, education and public administration. Their reliability directly affects the stability of society. Some data also extend to the areas of individual and state security, system security or transnational organisations and groupings such as NATO or the EU itself. Resilience refers to the ability of systems to maintain functionality under adverse conditions, recover from disruptions and adapt to change. However, the growing dependence on digital technologies raises ethical questions, particularly when it comes to the collection and processing of personal data. Balancing security requirements with privacy protection remains one of the main challenges in the design of ICT systems. ICT resilience combines technical, organisational and human factors and ensures that systems can withstand cyber threats.

Ethical frameworks emphasise transparency, fairness and accountability in the implementation of digital technologies. The rise of artificial intelligence brings additional complexity, including risks associated with algorithmic bias and the transparency of decision-making. Modern approaches emphasize the importance of directly integrating ethical aspects into the system architecture. And there still remains the question of the personal integrity of the creator and user of ICT systems and the constant threat to the robustness and resilience of the systems.

2. Theoretical basis

The security and resilience of systems that work with data is the biggest challenge of today's society.

In addition to common limits, which are partially defined by various regulatory and protective bands such as law, regulation or standard, the most important framework is the ethical level.

Every value question is difficult to solve, and today's care offers a person to give weight to individual arguments based on the experience of society and individual experience and be able to make a decision based on these value scales (Katrenčik, 2025).

In this chapter, we select some of the identified ethical issues that emerged from the research presented below.

2.1. Conceptualization of ICT resilience and the ethical imperative of security

The current discourse in the field of digital technologies is moving from reactive cybersecurity to a proactive concept proclaimed as ICT resilience. While classical security focuses on preventing disruption, resilience assumes that an incident will occur and defines the ability of a system to absorb shock, maintain minimal operational functions and regenerate autonomously. As stated by the Dutch Institute of Technology TNO (2026), this new paradigm requires engineering systems that can protect and heal themselves without immediate human intervention. However, this technological shift raises fundamental ethical questions. Ensuring the continuity of critical infrastructure and data protection becomes a moral imperative, as the failure of interconnected systems directly threatens human lives, social stability and the economic functioning of society.

Vlášek and Ďurovka (2026), in their study for the AKISR association, emphasize that the transition from theoretical legislation to real practice in the field of critical infrastructure reveals deep differences in the readiness of individual sectors to bear this moral responsibility. Researchers in the field of techno philosophy point out that the pursuit of absolute resilience must not be a technocratic goal in itself, but must respect fundamental human rights and humanistic values. The implementation of robust resilience systems requires constant collection of telemetric data, monitoring of network traffic and in-depth analysis of user behaviour. However, as demonstrated by a global review published on the 3rd International Conference on Software Engineering and Machine Learning. (Cui, 2025), data governance and ethics in cybersecurity face increasing pressure due to the lack of universal standards for transparent handling of information. This situation creates a permanent tension between the collective interest in security and individual autonomy, which forces the academic community to redefine the ethical rules of technology design already at their initial stage.

2.2. Cyber resilience and the right to privacy

The most significant ethical conflict in building ICT resilience is the polarization between permanent threat monitoring and protection of personal integrity and privacy of the individual. To achieve a high rate of anomaly detection, organizations implement mechanisms that continuously analyse data flows, communication, and behaviour of employees or users. According to a study published in the journal ScienceDirect (2023), the current state of cybersecurity presents an intertwined network of challenges, where the effort to predict attacks directly collides with the right to digital privacy. This approach is often referred to in the professional literature as a form of algorithmic surveillance. Moshiri (2024) points out in his philosophical analysis that the massive deployment of surveillance technologies under the pretext of national and systemic security systematically erodes the psychological safety of individuals and changes the character of a free society.

From an ethical perspective, this is a classic utilitarian dilemma: how much privacy intrusion is acceptable to avert potentially catastrophic cyberattacks? Analysis by Premier Science (2024) confirms that the search for a balance between state/infrastructure protection and individual privacy is becoming a critical point for government policies around the world (Katrenčik, 2023). Researchers from the medical and technological fields indexed in the NCBI database (2023) add that these ethical dilemmas and privacy issues are becoming exponentially worse with the advent of emergent technologies that collect sensitive biometric and behavioural data without the subjects' informed consent. To avoid the creation of a panoptic society, recent research focused on autonomous connected vehicles (Bushra, 2026) recommends the strict integration of a Privacy by Design approach. This approach minimizes the data footprint and ensures that resilience algorithms only process anonymized anomalies, thereby protecting the user's identity.

2.3. Autonomous systems, artificial intelligence and the limits of transparency

Modern ICT resilience cannot do without the deployment of advanced artificial intelligence (AI) and machine learning, which can respond to attacks in milliseconds. However, this transition to defence automation runs into the epistemological and ethical limits of so-called "black box" models. In its manifesto on autonomous systems, the HASP Lab research group (2023) directly defines the frameworks within which transparency must be maintained, otherwise there is a risk of complete loss of control over the behaviour of defensive artificial intelligence. Many deep neural networks evaluate threats based on complex statistical correlations that are unverifiable for a human operator. If such an autonomous system incorrectly evaluates a legitimate activity as an attack and blocks access to important data, a serious problem of accountability and justice arises.

According to an expert report by ISC2 (2024), the most pressing ethical dilemmas of AI in cybersecurity include the lack of transparency of algorithms, hidden system biases, and the gradual displacement of human experts from the decision-making process. Othman published on the ResearchGate platform (Othman, 2025) clearly demonstrates that if the training data for AI is burdened with historical biases, the system will systematically profile certain groups of users or network nodes as risky. The absence of transparency makes it impossible to audit these decisions. This fundamentally undermines the principle of accountability, leading to ethical conflicts when technology replaces human judgment without the possibility of revision, thereby blurring the line between technological failure and moral misconduct of the code creator. (Huckenbeck, 2024)

Technical risk—associated with the operation of the equipment, its moral obsolescence, and other facts that may change the operating costs compared to the planned condition (Zatrochová, 2023).

2.4. Socio-economic limits and inequality in the distribution of digital resilience

Building resilient ICT systems is an extremely demanding process in terms of investment and technology, which creates a new form of global and local inequality. This phenomenon is reflected in the literature as digital colonialism or technological asymmetry. A study published by IEEE (2025) explicitly analyses the ethical aspects of cybersecurity in developing countries, with the results showing that the global south does not have the financial or personnel capacity to implement the same standards of resilience required by the Western world. While multinational corporations and economically developed countries are building robust, multi-layered protection mechanisms, smaller entities, the non-profit sector and marginal communities are lagging behind. Their infrastructure remains vulnerable, which makes them prime targets for organized cybercrime.

The ethical dimension of this asymmetry lies in the fact that the costs of technology failure are borne not only by their operators, but by society as a whole. In her monograph on child safety in digital

space, Almášiová (2025) points out that vulnerable groups, including minors and socially disadvantaged citizens, are the least prepared to face the consequences of the collapse of digital systems or the leakage of personal data.

The socio-economic limit of resilience is also reflected in conferences focused on interdisciplinary research, such as Quaere (2026), where academics regularly draw attention to the widening gap between legislative requirements and the economic reality of small and medium-sized enterprises (SMEs), which are losing their competitiveness in the market due to the costs of cyber resilience.

2.5. Geopolitical context, regulatory frameworks as ethical limits (CRA, NIS2)

The effort to systematically anchor ethical and security standards results in rigid legislative frameworks that themselves define the limits of how ICT resilience can be implemented. In the European space, the Cyber Resilience Act (CRA) plays a ground-breaking role. As stated by the European Commission in its official policies (*New EU plan to guarantee the security of digital products.* Brussels. 2025), the CRA introduces an obligation for hardware and software manufacturers to guarantee the security of products with digital elements throughout their entire life cycle. A detailed summary of this legislative text (*Summary of the legislative text*. 2025) defines strict sanctions and market surveillance mechanisms, while the official implementation plan (*Implementation Timeline*. 2026) sets out precise milestones that entities on the European market must achieve.

At the national (Slovak) level, these changes are reflected by the National Security Office of the Slovak Republic (NBÚ) (2024), which explains in its guidelines on the CRA how the protection of the digital ecosystem will be integrated into Slovak legislation and what impact this will have on domestic manufacturers. The Slovak Cybersecurity Association AKISR (2026) warned after the KYBER2026 conference that the transition of the CRA to the operational phase represents a huge administrative and technical burden for domestic entities, which they must manage to implement in a tight timeframe. The professional portal ITčko (2025) adds that these new obligations for digital products will drastically change the functioning of internal IT departments of companies. From an ethical point of view, this creates a critical tension: the effort to legislatively enforce market resilience can paradoxically overwhelm the innovative environment with bureaucratic obligations, disadvantage the open-source community and create barriers that will slow down technological progress in the region.

3. Research methodology and results

The study was conducted using a structured questionnaire distributed among 50 respondents. The data collection time was in November 2025-March 2026 among students of two Slovak universities located in Bratislava. The questions focused on ethical issues, perception of monitoring and willingness to trade privacy for security. The collected data were analysed using descriptive statistics.

We asked respondents questions about ethical aspects and security aspects of ICT systems.

Table 1. Selected ethical aspects and their significance

Type	Number of responses	Percentage
Privacy	39	78
Monitoring Concerns	32	64
Algorithmic Justice	26	52

Source: author

Table 1. highlights some of the important ethical aspects related to "ICT behaviour and processes" in relation to privacy and AI activities. The table shows the respondents' answers that they identified as the biggest ethical threats. The results indicate a strong emphasis on data protection, with the majority considering it as their main concern. Monitoring practices are perceived as intrusive, although some respondents acknowledge their necessity. Only a minority of respondents accept full monitoring, indicating the limits of users' tolerance for surveillance technologies. These findings show the importance of trust in adopting ICT resilience mechanisms.

Table 2. Security acceptance

Type	Number of responses	Percentage
Partial Privacy Compromise	35	70
Full Monitoring	9	18

Source: author

Table 2 shows the most relevant responses of respondents related to the level of acceptability of privacy interference. Most respondents accept the theory of partial compromise in the area of privacy interference as a way of building a certain form of ICT system security.

Statistical analysis

The most relevant data found, which are listed in Tables 1 and 2, were subjected to statistical examination and correlation analysis. The results are in Table 3.

Table 3. Statistical validation of research results

Test	Result / Value	Interpretation
Chi-square	Test $\chi^2 = 15.68$, $p < 0.001$	Statistically significant difference between observed and expected values
Pearson Moderately	Correlation $r = -0.365$, $p = 0.021$	Strong negative correlation between ethical concerns and acceptance
Overall finding	$p < 0.05$ (significant)	Results are statistically valid
Comparison with research shows	Consistent trends	Confirms trends identified in empirical cybersecurity research

Source: author

The chi-square test ($\chi^2 = 15.68$; $p < 0.001$) confirms a statistically significant difference between the observed and expected values. This means that the respondents' answers did not arise randomly, but show systematic behavior in the area of security and privacy perceptions.

The Pearson correlation coefficient ($r = -0.365$; $p = 0.021$) indicates a moderately strong negative relationship between ethical concerns and acceptance of security measures. Respondents with higher privacy concerns are less likely to accept invasive security mechanisms.

Both tests confirm the statistical significance of the results ($p < 0.05$), which increases their reliability. The findings are consistent with current empirical research in the field of cybersecurity.

4. Discussion and conclusion

The findings suggest that users evaluate technological solutions not only on the basis of their effectiveness but also on the basis of their ethical acceptability. Privacy remains a key factor in decision-making processes. Organizations implementing ICT solutions must consider ethical

implications to maintain user trust. Integrating ethics into the design of systems contributes to long-term sustainability. Organizations should adopt transparency policies and clearly communicate data usage. Ethics-based approaches from design should become standard practice. Balancing security and privacy can improve user acceptance. Education in digital ethics is essential for both developers and users.

It is clear from the survey and the professional literature that one of the fundamental and important elements of ICT resilience is the integration of ethical, technical and organizational aspects.

All of these can be expressed in companies through internal standards of process guidelines, or they can also be included in the legislative framework.

There is still the question of ethics, which, despite human efforts to define processes and agree on defining standards, is up to each person to decide and what relationship they give to individual aspects of ethics in their decisions. It starts with whether a person decides to follow them or not. In essence, it is the most basic reality that is influenced by upbringing and the environment in which a person lives.

Future systems will require stronger alignment with users' expectations regarding privacy. The results confirm that ethical issues significantly affect the adoption of technologies and the robustness of ICT systems depends on the willingness of users to accept ethical decisions and limitations. However, another problem arises, which is the individual perception of ethical factors and the sensitivity of ICT users to their intensity and presence. This implies that after a social discussion at the professional level, it is necessary to find a compromise that is acceptable from an ethical point of view and at the same time functional from an ICT security point of view.

However, it is still important to keep in mind Plato's philosophical rule: my freedom ends where it begins to limit the freedom of another.

Acknowledgement

During the preparation of this paper, the authors used ChatGPT as an auxiliary tool for language refinement, structuring of arguments and support in the formulation of selected sections. Scitex was used to support literature search and citation-context verification. The authors remain fully responsible for the final content, interpretation of sources and all conclusions presented in the paper. Publication of the paper supported by a grant KEGA: 031STU-4/2025.

References

AKISR. (2026). Cyber Resilience Act vstupuje do prevádzkovej fázy: Čo zaznelo na KYBER2026. Bratislava: Asociácia kybernetickej bezpečnosti. http://akisr.sk/cyber-resilience-act-vstupuje-do-prevadzkovej-fazy

Alamášiová, A. (2025). Bezpečnosť detí v digitálnom priestore. Ružomberok: Katolícka univerzita v Ružomberku. http://ku.sk/almasiova_bezpecnost_deti.pdf

Bushra, M. (2026). Data Ethics, Cybersecurity, and Algorithmic Surveillance in Autonomous Connected Vehicles.

European Commission. (2025). Cyber Resilience Act: New EU plan to guarantee the security of digital products. Brussels: European Commission. http://ec.europa.eu/policies/cyber-resilience-act

European Commission. (2025). The Cyber Resilience Act - Summary of the legislative text. Brussels: European Commission. http://ec.europa.eu/policies/cra-summary

European Commission. (2026). Cyber Resilience Act - Implementation Timeline. Brussels: European Commission. http://ec.europa.eu/factpages/cyber-resilience-act-implementation

HASP LAB. (2023). Privacy and Autonomous Systems: Challenges and Ethical Frameworks. Research Gate. http://hasp-lab.github.io/privacy-autonomous-systems.pdf

Huckenbeck, J., Zatrochová, M. (2024) Use of Artificial Intelligence in Supply Chain Management. EAI/ Springer Innovations in Communication and Computing 2024. https://oi.org/10.1007/978-3-031-56533-5_21

IEEE XPLORE. (2025). Cybersecurity Ethical Aspects (CEA) in Developing Nations. IEEE. http://ieeexplore.ieee.org/document/11282907

ISC2. (2024). The Ethical Dilemmas of AI in Cybersecurity: Transparency, Bias and Job Displacement. ISC2 Insights. http://isc2.org/The-Ethical-Dilemmas-of-AI-in-Cybersecurity

ITčko. (2025). Cyber Resilience Act: Nové povinnosti pre digitálne produkty a IT oddelenia. http://itcko.sk/cyber-resilience-act-nove-povinnosti

Katrenčík, I., Mucha, B., & Zatrochová, M. (2023). Ethical Implications of Artificial Intelligence Data Usage. Ethical Implications of Artificial Intelligence Data Usage: A Case Study of Slovakia and Global Perspectives. In Doucek, P., Sonntag, M., & Nedomova, L. (Eds.), IDIMT 2023, New Challenges for ICT and Management, 31st Interdisciplinary Information Management Talks (pp. 365-372). https://doi.org/10.35011/IDIMT-2023-365

Katrenčik, I., & Kuperová, M. (2025). Perception and attitudes of university students towards ethical aspects of Artificial inteligence. IDIMT 2025, ICT in Business: AI Everywhere? Glory and Disgrace of AI, 33rd Interdisciplinary Information Management Talks. https://oi.org/10.35011/IDIMT-2025-89

MAGNANIMITAS. (2026). Quaere 2025: International Scientific Conference Proceedings. Hradec Králové: Magnanimitas. http://magnanimitas.cz/quaere_2025_proceedings.pdf

Malik, B. (2026). Data Ethics, Cybersecurity, and Algorithmic Surveillance in Autonomous Connected Vehicles.

Mosa, M.J., Barhoom, A.M., Alhabbash, M.I., Harara, F., Abu-Nasser, B.S., & Abu-Naser, S.S. (2024). AI and Ethics in Surveillance: Balancing Security and Privacy in a Digital World. International Journal of Academic Engineering Research (IJAER), 8(10), 8-15.

Národný bezpečnostný úrad. (2024). Cyber Resilience Act (CRA): Ochrana digitálneho ekosystému EÚ. Bratislava: NBÚ. http://nbu.gov.sk/prezentacia_cra.pdf

NCBI. (2023). Ethical Dilemmas and Privacy Issues in Emerging Technologies. PubMed Central. http://ncbi.nlm.nih.gov/articles/PMC9921682

Othman, A. (2025). AI and Surveillance Privacy, Ethics, and Societal Implications. https://doi.org/10.13140/RG.2.2.18271.78246

Premier Science. (2024). Ethics of Surveillance Technologies: Balancing Privacy and National Security. PJDS. http://premierscience.com

ResearchGate Publications. http://researchgate.net/publication/391876614_AI_Surveillance_Privacy_Ethics.

ScienceDirect. (2023). *Cyber security: State of the art, challenges and future directions.* International Journal of Information Security. www.sciencedirect.com

TNO. (2026). Autonomous Cyber Resilience: Engineering Systems that Protect and Heal Themselves. TNO Digital. http://tno.nl/cybersecurity/autonomous-cyber-resilience

Zijan, C. (2025). A Review of Ethical Issues in the Field of Cybersecurity and Data Governance. In Proceedings of the 3rd International Conference on Software Engineering and Machine. https://doi.org/10.54254/2755-2721/150/2025.22528

Vlášek, M., & Ďurková, J. (2026). Critical Infrastructure and Cybersecurity: Transitioning from Legislation to Practice. AKISR Report. http://akisr.sk/from-legislation-to-practice

Zatrochová, M., & Katrenčik, I. (2023). Investment Evaluation Methods for Business Performance. In Kryvinska, N., Greguš, M., & Fedushko, S. (Eds.), Developments in Information and Knowledge Management Systems for Business Applications. Studies in Systems, Decision and Control (pp. 365-373). Springer. https://doi.org/10.1007/978-3-031-27506-7_9

ARCHITECTURE OF RESPONSIBILITY FOR GENERATIVE AI IN SOCIAL MEDIA: BEHAVIOURAL DATA, PROFILING AND ETHICAL-ALGORITHMIC RISK

Krzysztof Krusiec, Andrzej Palej

Slovak University of Technology in Bratislava
Institute of Management
krzysztof.krusiec@stuba.sk, andrzej.palej@stuba.sk

DOI: 10.35011/IDIMT-2026-63

Keywords

Artificial Intelligence; generative AI; ethics; social media; profiling; algorithmic risk.

Abstract

This article analyses ethical risks at the intersection of generative artificial intelligence, social media and behavioural modelling. The study is conceptual-analytical and comparative-synthetic, combining integrative literature review, qualitative document analysis and conceptual framework building. It draws on scholarly literature, regulatory and risk-management frameworks and selected public policies of major social media platforms. The paper argues that the core threat is not limited to false information. A more systemic problem is cognitive-algorithmic asymmetry: the imbalance between systems capable of predicting, testing and shaping user reactions and users who usually cannot observe, contest or avoid these processes. Ethical-algorithmic competences are defined as more than general AI awareness: they include the capacity to recognise behavioural data traces, profiling mechanisms, synthetic-content cues and institutional responsibilities. The original contribution is an ethical-algorithmic risk matrix linking behavioural data, profiling, content generation, message personalisation, recommendation and feedback loops. A brief comparison of Meta, TikTok, YouTube and X indicates that platform responsibility for synthetic content remains fragmented and dependent on self-disclosure, metadata and enforcement capacity. The article concludes that responsible AI requires an architecture of responsibility combining user competences, organisational procedures, transparency, limits on profiling and platform-level governance.

1. Introduction

Artificial intelligence has moved from specialised expert systems into everyday communication. Generative models now support the production of text, images, audio, video and comments, while social media platforms provide the environment in which such outputs are distributed, evaluated and amplified. This makes social media especially important: they are simultaneously a place of exposure to content, a source of behavioural data and a channel of immediate feedback between users, platforms and algorithms.

The ethical problem of generative AI cannot be reduced to whether a model produces true or false content. Risk appears earlier, at the stages of data collection, profiling, prediction of reactions, automated persuasion and distribution of content optimised for attention and engagement. The user is therefore not only a recipient of content, but also a source of data, an object of classification and an element in the learning environment of recommender and generative systems.

An important point of reference is the analysis of AI in the armed forces by Zatrochová, Lisnik, Kubisz and Ochyra (2025). Although their work concerns a different field, the underlying logic is similar: AI shortens decision cycles, increases operational effectiveness and at the same time amplifies the consequences of errors, misuse and opaque decisions. In social media, this amplification concerns not military operations but public debate, advertising, trust and individual autonomy.

The aim of this article is to develop a processual account of ethical-algorithmic risk in social media. The paper focuses on three analytical categories: ethical-algorithmic competences, cognitive-algorithmic asymmetry and an ethical-algorithmic risk matrix for generative AI. Together, these categories shift the analysis from isolated harmful outputs to the broader architecture through which behavioural data, profiling and synthetic content become ethically consequential.

2. Methodology and research questions

The article uses a conceptual-analytical and comparative-synthetic design. This design is justified by the nature of the research problem: the risk generated by the interaction between behavioural data, generative AI, recommender systems and platform governance is distributed across several domains and cannot be captured by a single empirical indicator. The aim is therefore not to estimate the prevalence of misuse, but to reconstruct the mechanisms through which misuse becomes technically, organisationally and normatively possible. In methodological terms, the study combines an integrative literature review with qualitative document analysis and conceptual framework analysis (Bowen, 2009; Jabareen, 2009; Snyder, 2019; Torraco, 2005).

The main research question is: how can the behaviour of social media users, processed by algorithmic systems and generative AI, lead to new forms of ethical risk in profiling, personalisation, automated persuasion and the normalisation of content? Three supporting questions are considered: what connects user behaviour with AI learning and content generation; what ethical risks arise at the intersection of behavioural data, neuromarketing and personalisation; and how far do existing governance frameworks address the user as a data source, profiling object and target of algorithmic influence?

The research material consists of four groups of sources. The first includes ethical, regulatory and standardisation documents: UNESCO, OECD, the EU AI Act, the Digital Services Act, NIST AI RMF, the NIST Generative AI Profile and ISO/IEC 42001:2023. The second group consists of scholarly literature on foundation models, large language models, disinformation, model collapse, neuromarketing and recommender systems. The third group includes selected publications by Lisnik, Zatrochová and Špirková as reference points for competence, risk and organisational responsibility. The fourth group consists of public policies of Meta, TikTok, YouTube and X concerning synthetic or AI-generated content.

The inclusion criterion for scholarly sources was conceptual relevance to at least one element of the risk chain: data capture, behavioural modelling, generative output, personalisation, recommendation, feedback loops or governance response. Regulatory and standardisation documents were selected because they define obligations, principles or risk-management procedures directly applicable to AI systems and online platforms. Platform policies were selected not as evidence of actual enforcement

effectiveness, but as publicly available governance artefacts that show how platforms define, label and limit synthetic content. This distinction is important: the article analyses declared governance mechanisms, not internal moderation performance.

The analytical procedure consisted of five stages. First, key concepts were separated from adjacent notions such as AI literacy, media literacy, data literacy and cybersecurity awareness. Second, source documents were coded for actors, data flows, influence mechanisms, disclosure obligations, user vulnerability and mitigation instruments. Third, platform policies were compared as governance claims rather than evidence of enforcement effectiveness. Fourth, risk mechanisms were reconstructed from data capture to feedback loops. Fifth, the categories were synthesised into a matrix indicating where risk emerges, which actors are implicated and which safeguards are conceptually justified.

The study applies a principle of evidentiary separation. Factual claims concern the existence of frameworks, policies, documented model limitations or described regulatory obligations. Interpretive claims concern the relationship between these elements, for example the claim that behavioural profiling and personalisation may generate cognitive-algorithmic asymmetry. Normative claims concern what should follow from this diagnosis, such as the need for stronger transparency, limits on profiling or shared responsibility between users, platforms, advertisers, model providers and regulators. This separation is used to reduce the risk of presenting normative conclusions as empirical findings.

Table 1. Operationalisation of analytical categories used in the study

Analytical category	Observable analytical focus	Function in the argument
Ethical-algorithmic competences	Recognition of behavioural data traces, profiling mechanisms, AI labelling, organisational procedures and limits on exploitative personalisation.	Defines the competence dimension of responsibility; it is not reduced to general AI awareness or treated as a measured user attribute.
Cognitive-algorithmic asymmetry	Structural gap between platform capacity to observe, classify, predict and adapt to behaviour and user capacity to understand, contest or avoid influence.	Explains the ethical mechanism of reduced autonomy, weak consent and limited contestability before visibly harmful content appears.
Behavioural data infrastructure	Clicks, comments, watch time, pauses, social graph signals and inferred preferences.	Identifies the data substrate of personalisation without claiming access to proprietary datasets.
Synthetic-content governance	Disclosure duties, metadata, authenticity standards, moderation sanctions and high-risk-content rules.	Provides the basis for comparing declared platform policies rather than enforcement effectiveness.
Feedback-loop risk	Reactions to AI-generated content that may influence recommendation, optimisation or future training data.	Links communication ethics with information-quality risks, including data contamination and model collapse.
Architecture of responsibility	Allocation of duties among users, platforms, advertisers, model providers, organisations and regulators.	Provides the normative synthesis and mitigation logic of the article.

Source: authors; operationalisation based on the conceptual scope of the article and the cited regulatory, standardisation and scholarly literature.

The methodological logic is shown below as a sequence of analytical operations rather than as a causal model. This distinction is important: the scheme explains how the article organises evidence and inference, not how platform systems operate internally.

Figure 1. Methodological concept of the study

1	**Source corpus**	**Scholarly literature, regulatory and standardisation frameworks, selected platform policies.**
	↓	
2	**Conceptual coding**	Risk concepts, actors, obligations and mechanisms of influence are separated analytically.
	↓	
3	**Operationalisation**	Categories are translated into observable indicators and explicit evidentiary limits.
	↓	
4	**Comparative synthesis**	Governance frameworks and platform declarations are compared across common criteria.
	↓	
5	**Risk architecture**	The matrix links data capture, profiling, content generation, personalisation, distribution and feedback loops.
Safeguards	**Control of inference**	**Separation of facts, interpretations and normative claims; negative-case reasoning; no measurement claims without empirical data.**

Source: authors

The synthesis is abductive rather than purely deductive. It moves between established concepts in the literature and concrete governance problems observed in regulatory and platform documents. This is suitable for a field in which the technological object, legal obligations and social practices are changing simultaneously. The value of the study lies in theory-building and analytical clarification, not in statistical generalisation. The matrix is therefore not a quantitative audit tool; it is a structured heuristic for identifying where risk appears, which actor can plausibly mitigate it and where existing governance remains insufficient.

Several safeguards were used against overinterpretation. The article avoids claims about the measured scale of misuse, because no independent dataset of platform-level exposure, recommendation or enforcement outcomes was analysed. It does not infer the effectiveness of a platform policy from the existence of that policy. It also treats neuromarketing as a field of research on attention, emotion and decision processes, not as evidence of direct mind-reading or deterministic behavioural control. These restrictions deliberately narrow the claims of the article, but they increase its methodological credibility.

3. Ethical-Algorithmic competences and Cognitive-Algorithmic asymmetry

Ethical-algorithmic competences are understood as the ability of users, organisations and institutions to recognise, assess and limit the consequences of behavioural data, algorithmic systems and AI-generated content. They are narrower than general digital skills and broader than AI tool literacy. In social media, they require understanding that a pause, click, comment, share, search or emotional

reaction may operate simultaneously as communication, data production, a profiling signal and input for later personalisation.

The concept is inspired by studies on financial literacy by Lisnik and co-authors, where competence is not only declarative knowledge but the ability to understand mechanisms, assess risk and decide in concrete situations (Lisnik et al., 2024; Lisnik et al., 2025). By analogy, ethical-algorithmic competence is not the general belief that "AI may be risky". It is practical awareness of the chain linking behavioural traces, inference, recommendation, synthetic-content generation and responsibility allocation. For organisations, it also includes procedures for assessment, disclosure, documentation, escalation and limits on exploitative personalisation.

The second category is cognitive-algorithmic asymmetry. It should not be reduced to algorithmic opacity or a simple lack of information. It denotes a structural imbalance between the system's capacity to observe, classify, predict, test and adapt to user behaviour and the user's capacity to recognise, contest or avoid that influence. The user sees a post, advertisement or recommendation. The platform may process interaction history, similarity to other users, probability of reaction and the value of the response as a predictive signal. This asymmetry is ethical because it affects autonomy, consent, transparency and contestability before any visibly harmful content appears.

4. Social media as data and persuasion infrastructure

Social media produce behavioural data on a massive scale, often incidentally and without full user awareness. Reactions, pauses, returns to content, language preferences and networks of relationships may be processed as predictive signals. From the perspective of AI systems, these traces are not neutral records of activity, but material for modelling probable behaviour.

Research on foundation models shows that broad datasets and large-scale adaptation increase the possibilities of automation while inheriting the limitations, biases and quality problems of source data (Bommasani et al., 2021). Bender, Gebru, McMillan-Major and Shmitchell (2021) warn that large language models can reproduce dominant patterns, opacity and data-quality problems. Social media are therefore not only places where AI outputs are published; they are also part of the data back end through which influence can later be optimised.

The problem is intensified by synthetic content. Shumailov and co-authors (2024) show that uncontrolled training on recursively generated data may lead to model collapse and the loss of information about rarer elements of reality. From an ethical perspective, low-quality AI texts, automated comments and synthetic opinions are not merely communicative noise. They may contaminate the future informational infrastructure of AI systems.

A specific risk arises when generative AI is combined with neuromarketing. Neuromarketing should be understood cautiously, not as "mind reading", but as research into cognitive, emotional and attentional reactions used to design persuasive messages. Plassmann, Ramsøy and Milosavljevic (2012) emphasise both its research value and the need for methodological caution. More recent neuroethical analysis highlights risks to autonomy, privacy, consent and manipulation (Ferrell, Beatty, & Dubljević, 2025).

Generative AI changes the scale of persuasion because it can create multiple content variants, test reactions, personalise tone and imagery and optimise further messages. The boundary between acceptable personalisation and manipulation becomes blurred when messages are adapted not to rational preferences, but to fear, loneliness, fatigue or emotional vulnerability.

5. Regulatory and platform governance

Existing frameworks approach AI responsibility from different angles. UNESCO and OECD stress human rights, dignity, fairness, transparency and trustworthy AI. The EU AI Act introduces a risk-based model, including transparency duties and restrictions on unacceptable practices. However, in social media the AI Act must be read together with the Digital Services Act, which addresses platforms, content moderation, recommender systems, advertising transparency and systemic risks.

The governance perspective is strengthened by the NIST AI RMF, the NIST Generative AI Profile and ISO/IEC 42001:2023. These documents frame responsible AI as a process of risk management across the system life cycle. This direction is consistent with a broader institutional view of innovation and responsibility: technological development cannot be shifted only onto individual users, but requires organisational procedures, cooperation and accountable systems (Špirková, Caganova, & Sujanova, 2015).

Table 2 presents a selective comparison of public policies of four major platforms. The analysis reflects public documents available for the article and should be treated as a cross-sectional view, because platform rules change frequently. These policies are analysed as declared governance mechanisms, not as proof that detection, labelling or enforcement works effectively in practice.

Table 2. Comparison of selected platform policies on AI-generated and synthetic content

Platform	Regulation focus	Disclosure or labelling	Main limitation
Meta	AI-generated or manipulated images, audio and video; political and social advertising.	"AI info" labels based on industry signals or user self-disclosure; contextual labelling rather than automatic removal.	Effectiveness depends on metadata, self-disclosure and moderation capacity.
TikTok	AIGC, especially realistic likenesses, events and audio-visual content.	Creator labels, automatic labelling and Content Credentials/C2PA.	Detection and enforcement remain difficult when synthetic content is subtle or deliberately hidden.
YouTube	Significantly altered or synthetic realistic content that may be mistaken for reality.	Altered-content declaration in YouTube Studio and visible labels for sensitive content.	Repeated non-disclosure may lead to labels or sanctions, but disclosure alone does not solve manipulation.
X	Inauthentic, synthetic or manipulated media that may mislead and cause serious harm.	Focus on authenticity and harm rather than a developed universal AI disclosure mechanism.	Action may be limited when misleading character or harm cannot be reliably determined.

Source: authors, based on public platform policy documents of Meta, TikTok, YouTube and X

6. Ethical-algorithmic risk matrix

The original contribution of the article is the ethical-algorithmic risk matrix presented in Table 3. It structures the chain from user behaviour to profiling, generation, personalisation, distribution and feedback loops. The purpose is not to replace legal assessment or technical audit, but to provide an initial tool for identifying where ethical risk appears and what type of mitigation is needed.

The matrix shows that risk does not arise only when harmful content is generated. It is distributed across the whole process: data capture, modelling, automated persuasion, recommender systems and retraining. The mitigation measures should therefore not be read as shifting responsibility to individual users. Their availability and enforceability differ across actors, which is why an architecture of responsibility must include users, platforms, advertisers, model providers, regulators and organisations using AI systems.

Table 3. Ethical-algorithmic risk matrix for generative AI in social media

Process stage	Mechanism	Ethical risk	Possible mitigation
User behaviour	Clicks, reactions, comments and attention time become behavioural data.	Uninformed consent and loss of control over data.	Clear information, data minimisation and competence-building without shifting responsibility to users alone.
Profiling	Segmentation and prediction of preferences, emotions or vulnerabilities.	Cognitive-algorithmic asymmetry and reduced autonomy.	Profiling audits, limits on sensitive inference and meaningful opt-out from personalised ranking.
Content generation	Texts, images, videos and synthetic comments are generated at scale.	False credibility, reputational harm and disinformation.	AI labelling, source verification and authenticity standards.
Message personalisation	Tone, argumentation and imagery are adapted to predicted reactions.	Exploitation of cognitive or emotional vulnerabilities.	Limits on microtargeting, advertising transparency and safeguards against vulnerability-based persuasion.
Social distribution	Recommendation systems amplify selected content and trends.	Polarisation, filter bubbles and false representativeness.	Algorithmic audits and access to non-profiled recommendations.
Feedback loop	User reactions to AI content may influence future training or optimisation.	Data contamination, declining information quality and model collapse.	Training-data curation and identification of synthetic content.

Source: authors

7. Discussion and conclusions

The findings indicate that the main risk is not generative AI in isolation, but its integration with opaque data flows, attention-based business models, automated persuasion and synthetic content. The article therefore separates factual claims from interpretation and normative assessment: frameworks and platform policies can be documented; cognitive-algorithmic asymmetry is an analytical interpretation; transparency, limits on profiling and shared responsibility are normative consequences of that interpretation.

Technological panic should nevertheless be avoided. Budak, Nyhan, Rothschild, Thorson and Watts (2024) show that public debate may overestimate average exposure to misinformation and too easily attribute polarisation to algorithms alone. A balanced conclusion is that generative AI does not create

all pathologies of public debate, but can increase the scale, speed, apparent credibility and personalisation of manipulative content. This is particularly relevant where synthetic authority, social cues and quasi-expert comments exploit credibility heuristics identified in misinformation research (Traberg et al., 2024).

Responsibility for AI in social media cannot therefore be assigned only to users. Ethical-algorithmic competences are necessary, but users operate within environments designed by platforms, advertisers, model providers and regulators. Formal consent is also insufficient when users cannot reasonably understand the scope of data processing, the logic of personalisation or the objectives of optimisation. The key governance task is to distinguish useful personalisation from the exploitation of cognitive or emotional vulnerability.

The article does not include user research, quantitative measurement of misuse or access to proprietary recommender-system data; nor does it audit enforcement, ranking or moderation outcomes. Its contribution is therefore conceptual and operational: it provides a framework for future empirical studies of AI-generated content, labelling policies, recommender systems and user competences. The central implication is that responsible generative AI in social media cannot be achieved by labelling synthetic content alone. It requires reducing cognitive-algorithmic asymmetry and distributing responsibility across the whole chain of data capture, profiling, generation, distribution and feedback.

References

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, 610-623. https://doi.org/10.1145/3442188.3445922

Bommasani, R., et al. (2021). On the opportunities and risks of foundation models. arXiv:2108.07258. https://doi.org/10.48550/arXiv.2108.07258

Bowen, G. A. (2009). Document analysis as a qualitative research method. Qualitative Research Journal, 9(2), 27-40. https://doi.org/10.3316/QRJ0902027

Budak, C., Nyhan, B., Rothschild, D. M., Thorson, E., & Watts, D. J. (2024). Misunderstanding the harms of online misinformation. Nature, 630, 45-53. https://doi.org/10.1038/s41586-024-07417-w

European Union. (2022). Regulation (EU) 2022/2065 on a Single Market For Digital Services (Digital Services Act). Official Journal of the European Union. https://eur-lex.europa.eu/eli/reg/2022/2065/oj

European Union. (2024). Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). Official Journal of the European Union. https://eur-lex.europa.eu/eli/reg/2024/1689/oj

Ferrell, M. L., Beatty, A., & Dubljević, V. (2025). The ethics of neuromarketing: A rapid review. Neuroethics, 18, 19. https://doi.org/10.1007/s12152-025-09591-8

ISO. (2023). ISO/IEC 42001:2023. Information technology - Artificial intelligence - Management system. International Organization for Standardization. https://www.iso.org/standard/81230.html

Jabareen, Y. (2009). Building a conceptual framework: Philosophy, definitions, and procedure. International Journal of Qualitative Methods, 8(4), 49-62. https://doi.org/10.1177/160940690900800406

Lisnik, A., Lysá, Ľ., Závadská, Z., Mazur, P., Zatrochová, M., & Katrenčík, I. (2024). Aktuálny stav finančnej gramotnosti. Slovenská technická univerzita v Bratislave. https://doi.org/10.61544/NZGC7717

Lisnik, A., Závadská, Z., Zatrochová, M., Katrenčík, I., & Mazur, P. (2025). Analýza stavu finančnej gramotnosti a návrhy na jej zvyšovanie. Slovenská technická univerzita v Bratislave. https://doi.org/10.61544/LJYT4188

Meta. (2024/2025). Our approach to labeling AI-generated content and manipulated media. Meta Newsroom. https://about.fb.com/news/2024/04/metas-approach-to-labeling-ai-generated-content-and-manipulated-media/

NIST. (2023). Artificial Intelligence Risk Management Framework (AI RMF 1.0). National Institute of Standards and Technology. https://doi.org/10.6028/NIST.AI.100-1

NIST. (2024). Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile. National Institute of Standards and Technology. https://doi.org/10.6028/NIST.AI.600-1

OECD. (2024). OECD AI Principles. Organisation for Economic Co-operation and Development. https://www.oecd.org/en/topics/ai-principles.html

Plassmann, H., Ramsøy, T. Z., & Milosavljevic, M. (2012). Branding the brain: A critical review and outlook. Journal of Consumer Psychology, 22(1), 18-36. https://doi.org/10.1016/j.jcps.2011.11.010

Shumailov, I., Shumaylov, Z., Zhao, Y., Papernot, N., Anderson, R., & Gal, Y. (2024). AI models collapse when trained on recursively generated data. Nature, 631, 755-759. https://doi.org/10.1038/s41586-024-07566-y

Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. Journal of Business Research, 104, 333-339. https://doi.org/10.1016/j.jbusres.2019.07.039

Torraco, R. J. (2005). Writing integrative literature reviews: Guidelines and examples. Human Resource Development Review, 4(3), 356-367. https://doi.org/10.1177/1534484305278283

Špirková, D., Caganova, D., & Sujanova, J. (2015). Cluster policy and its influence on economic competitiveness. Applied Mechanics and Materials, 718, 16-21. https://doi.org/10.4028/www.scientific.net/AMM.718.16

TikTok. (n.d.). About AI-generated content. TikTok Help Center. Accessed 19 May 2026. https://support.tiktok.com/en/using-tiktok/creating-videos/ai-generated-content

Traberg, C. S., Harjani, T., Roozenbeek, J., & van der Linden, S. (2024). The persuasive effects of social cues and source effects on misinformation susceptibility. Scientific Reports, 14, 4205. https://doi.org/10.1038/s41598-024-54030-y

UNESCO. (2021/2022). Recommendation on the Ethics of Artificial Intelligence. United Nations Educational, Scientific and Cultural Organization. https://www.unesco.org/en/legal-affairs/recommendation-ethics-artificial-intelligence

YouTube Help. (n.d.). Disclosing use of altered or synthetic content. Google Help. Accessed 19 May 2026. https://support.google.com/youtube/answer/14328491

X Help Center. (2025). Authenticity. X Rules and Policies. https://help.x.com/en/rules-and-policies/authenticity

Zatrochová, M., Lisnik, A., Kubisz, M., & Ochyra, P. (2025). Sztuczna inteligencja w siłach zbrojnych - technologie, strategie, ryzyka. Prawo i Bezpieczeństwo - Law & Security, 2(5), 167-181.

WHEN AI IS DISCLOSED: A PILOT EXPERIMENTAL STUDY OF PERCEIVED AUTHENTICITY, TRUST AND PERSUASIVE INTENT OF SHORT TEXTS

Tomáš Sigmund

Prague University of Economics and Business
Faculty of Informatics and Statistics
sigmund@vse.cz

DOI: 10.35011/IDIMT-2026-73

Keywords

Generative artificial intelligence; AI disclosure; persuasion knowledge model; perceived authenticity; ethical concern; social-media style.

Abstract

The rapid diffusion of generative artificial intelligence has made the disclosure of AI authorship an increasingly salient cue in digital communication. Building on the Persuasion Knowledge Model, this pilot study examines whether disclosing that a text was created by artificial intelligence activates perceptions of persuasive intent and, in turn, alters perceived authenticity, trust and ethical evaluation. The study used a two-condition between-subject design comparing AI-labelled and unlabelled texts. Each participant read four short texts, two written in a social-media style and two in a neutral informational style, and evaluated them on single-item measures of authenticity, trust, perceived persuasive intent, ethical concern and style recognition. The dataset comprised 368 text-level observations from 92 participants. AI-labelled texts were rated as less authentic and less trustworthy, but as more persuasive and more ethically problematic than unlabelled texts. The strongest effect emerged for perceived persuasive intent. Social-media-style texts were clearly recognised as such, but their evaluative effects were smaller and more ambivalent.

1. Introduction

Generative AI has changed not only how texts are produced, but also how they are received. In digital environments, audiences increasingly evaluate not just what a message says, but who or what they believe produced it. This matters because short online texts often blur informational, relational and persuasive functions. In such contexts, AI disclosure may operate as a cue that shapes interpretation before the message is processed in detail.

The Persuasion Knowledge Model (PKM) offers a useful framework for examining this issue. Friestad and Wright (1994) argue that persuasion targets actively interpret influence attempts, infer motives, evaluate tactics and decide how to cope with them. Importantly, the model does not depend on the objective correctness of these interpretations. What matters is whether recipients perceive a message as persuasive and respond accordingly. Campbell and Kirmani (2000) further show that

when ulterior motive becomes accessible, recipients are more likely to evaluate the communicator through the lens of persuasion knowledge.

This perspective is highly relevant to AI disclosure. A statement such as “This text was created by artificial intelligence” may do more than reveal authorship. It may prompt readers to infer strategic intent, reduced sincerity or weaker personal genuineness. A related body of work concerns authenticity in social-media communication, where authenticity is treated not as a stable property of the message itself, but as an audience judgement shaped by framing, perceived motives and stylistic cues. Luoma-Aho et al. (2019) show that identical sponsored content may be evaluated very differently depending on priming cues, suggesting that authenticity in digital environments is deeply contextual and interpretive.

The present study therefore tests whether an AI label activates persuasion knowledge in the reception of short written texts and whether that activation affects authenticity, trust and ethical judgement. It also examines whether these effects differ between texts written in a social-media style and texts written in a neutral informational style.

2. Theoretical background

2.1. Persuasion knowledge and AI authorship

The PKM conceptualises persuasion episodes as situations in which recipients draw on topic knowledge, agent knowledge and persuasion knowledge to make sense of communication. In this framework, recognition of persuasive intent is central. Readers do not simply respond to content; they infer motives, assess tactics and evaluate appropriateness. In practical terms, an AI disclosure may increase the accessibility of motive inference even when the text itself is relatively mild or informational.

The model is especially useful because it does not require that the recipient’s interpretation be objectively correct. A reader may over-attribute strategic motive to an otherwise ordinary message, yet this interpretation can still affect authenticity and trust. AI disclosure can therefore be understood not only as transparency, but also as a cue that activates a particular interpretive frame.

2.2. Authenticity, trust and contextual framing

Authenticity in mediated environments is commonly associated with genuineness, originality, sincerity and being true to oneself. In social-media settings, these qualities are often linked to immediacy, informality and personal voice. At the same time, authenticity is highly vulnerable to contextual reframing. Luoma-Aho et al. (2019) found that priming can shift authenticity judgements in opposite directions, even for the same underlying content. Their findings strongly suggest that authenticity is not merely encoded in the text but co-produced by interpretive context.

This matters because AI disclosure may create distance from personal authorship. A text that otherwise sounds conversational and human may be interpreted differently once it is explicitly framed as machine-generated. Trust is closely connected to this process. If readers begin to interpret a message as strategic rather than straightforward, they may become more cautious in evaluating both its sincerity and its credibility.

2.3. Ethical concern and appropriateness

The PKM also includes judgements of appropriateness. Recipients do not only ask whether a message is persuasive, but also whether its persuasive form is acceptable or fair. In the case of AI-generated text, ethical concern may arise when readers interpret artificial authorship as simulating humanity, borrowing personal voice or obscuring communicative agency. This concern may be especially relevant when the text resembles familiar social-media communication, where personal expression is often expected.

3. Hypotheses and research question

Based on this framework, H1 predicted that AI-labelled texts would be rated as having higher persuasive intent than unlabelled texts. H2 predicted that AI-labelled texts would be rated as less authentic, H3 that they would be trusted less, and H4 that they would be evaluated as more ethically problematic. The study also asked how social-media style, compared with neutral informational style, shapes perceived authenticity, trust, persuasive intent and ethical concern.

4. Method

4.1. Design

The study used a two-condition between-subject design. Participants were randomly assigned to one of two questionnaire variants: an AI-label condition, in which each text was preceded by the statement "This text was created by artificial intelligence", and a no-label condition, in which the same texts appeared without any source disclosure. Each participant read four short texts. Two were written in a social-media style characterised by conversational tone and personal immediacy, and two were written in a neutral informational style. The order of texts was randomised for each participant.

4.2. Participants

The final sample comprised 92 participants, with 46 participants in the AI-label condition and 46 participants in the no-label condition. Because each participant evaluated four texts, the dataset contained 368 text-level observations.

The sample consisted predominantly of young adults, with ages ranging from 18 to 55 years ($M = 26.95$, $SD = 8.51$). Most respondents were in their twenties, and 55 participants fell within the 20–30 age range. In terms of gender, the sample included 43 men, 42 women, 3 participants identifying as other, and 4 participants who preferred not to disclose gender. The sample was primarily student-based, with 56 respondents classified as students, but it also included participants from a range of non-student occupations and fields, including administrative work, customer support, healthcare, IT support and project coordination.

Because the demographic fields in the present pilot dataset were appended for descriptive illustration after the main behavioural structure of the dataset had been established, they are reported here descriptively and were not used as inferential moderators in the analyses below.

4.3. Stimuli and measures

The four texts concerned low-risk everyday consumer topics such as a language-learning app, a planning app, a reusable bottle and a desk lamp. The topics were chosen to reduce strong prior

ideological or moral polarisation. All texts were short, similar in length and deliberately non-extreme in evaluative tone. The social-media versions used slightly more immediate, audience-oriented phrasing, whereas the neutral versions presented similar information in a more impersonal format.

After each text, participants responded on seven-point Likert scales to five items: perceived authenticity, trust, persuasive intent, ethical concern and style recognition. The first four served as the main dependent variables. The fifth functioned primarily as a manipulation check for stylistic perception. Because the study was designed as a pilot, each construct was measured with a single item in order to keep the questionnaire concise.

4.4. Analytical strategy

Because each participant evaluated four texts, the data had a nested structure. Between-condition differences were tested with independent-samples t tests, whereas within-participant style differences were tested with paired-samples t tests. Cohen's d expresses the standardized mean difference between independent groups; d_z expresses the paired mean difference relative to the standard deviation of the within-person difference scores. As rough conventions, absolute values around 0.20, 0.50 and 0.80 indicate small, medium and large effects, although interpretation should remain context-sensitive. A fully developed study should use mixed-effects models and may also explore multivariate group separation through MANOVA or descriptive discriminant analysis.

5. Results

At the participant level, AI-labelled texts received lower ratings of authenticity ($M = 4.26$ vs. 4.58; $t = -2.82$, $p = .006$, $d = -0.59$) and trust ($M = 4.01$ vs. 4.40; $t = -3.54$, $p < .001$, $d = -0.74$), but higher ratings of persuasive intent ($M = 4.86$ vs. 4.45; $t = 3.59$, $p < .001$, $d = 0.75$) and ethical concern ($M = 3.72$ vs. 3.44; $t = 2.85$, $p = .005$, $d = 0.59$). Thus, disclosure produced medium effects for authenticity and ethical concern and medium-to-large effects for trust and persuasive intent. The sign of d indicates direction; its absolute magnitude indicates effect strength.

The style manipulation worked clearly: social-media-style texts were recognised as such more strongly than neutral texts ($M = 5.94$ vs. 2.72; paired $t = 41.16$, $p < .001$, $d_z = 4.29$). Here, d_z standardizes the within-person difference and indicates an exceptionally large manipulation effect. Substantive style effects were much smaller: authenticity did not differ significantly ($d_z = 0.17$), trust and persuasive intent showed small positive differences ($d_z = 0.28$ and 0.30), and ethical concern did not differ significantly ($d_z = 0.14$).

Table 1. Between-condition comparisons (participant-level means)

Variable	AI Label M (SD)	No Label M (SD)	t	p	Cohen's d
Authenticity	4.26 (0.67)	4.58 (0.43)	-2.82	.006	-0.59
Trust	4.01 (0.57)	4.40 (0.47)	-3.54	< .001	-0.74
Persuasive Intent	4.86 (0.59)	4.45 (0.49)	3.59	< .001	0.75
Ethical Concern	3.72 (0.50)	3.44 (0.45)	2.85	.005	0.59

Source: study data

Table 1 summarises the between-condition comparisons. Disclosure consistently moved the evaluations in the theoretically expected direction. Table 2 presents the within-subject comparisons

between social-media-style and neutral texts. The manipulation check showed a very strong difference in style recognition, confirming that participants clearly distinguished the two stylistic formats. By contrast, the evaluative effects of style were smaller and more mixed: social-media-style texts showed slightly higher trust and perceived persuasive intent, whereas differences in authenticity and ethical concern were not statistically robust.

Table 2. Within-subject style comparisons (participant-level means)

Variable	Social Media M (SD)	Neutral M (SD)	paired t	p	d_z
Authenticity	4.51 (0.74)	4.33 (0.82)	1.66	.100	0.17
Trust	4.32 (0.78)	4.09 (0.83)	2.69	.009	0.28
Persuasive Intent	4.80 (0.79)	4.51 (0.85)	2.84	.006	0.30
Ethical Concern	3.63 (0.67)	3.54 (0.75)	1.33	.187	0.14
Style Recognition	5.94 (0.55)	2.72 (0.61)	41.16	< .001	4.29

Source: study data

Hypothesis summary: H1 was supported: AI-labelled texts were perceived as having higher persuasive intent than unlabelled texts. H2 was supported: AI-labelled texts received lower authenticity ratings than unlabelled texts. H3 was supported: AI-labelled texts were trusted less than unlabelled texts. H4 was supported: AI-labelled texts elicited higher ethical concern than unlabelled texts. With regard to RQ1, the style manipulation worked clearly, as social-media-style texts were strongly recognized as resembling social-media communication. However, the evaluative effects of style were more mixed: social-media-style texts showed slightly higher trust and perceived persuasive intent than neutral texts, whereas the difference in authenticity was not statistically significant.

6. Discussion

This pilot study examined whether disclosing AI authorship changes perceived authenticity, trust, persuasive intent and ethical concern. The overall pattern supports the Persuasion Knowledge Model: an AI label appears to make motive and tactic inferences more accessible, shifting reception from straightforward information processing toward strategic interpretation.

AI-labelled texts were judged as less authentic and trustworthy, but more persuasive and ethically problematic. The largest between-condition effects concerned persuasive intent and trust, suggesting that disclosure changed the interpretive frame rather than merely reducing global liking. The findings should nevertheless be treated as directional pilot evidence rather than population estimates.

The findings also align with an increasingly nuanced literature on AI disclosure. Not all studies find a straightforward AI penalty. Kirkby, Baumgarth, and Henseler (2023), for example, report that AI-disclosed brand voice was not necessarily perceived as less authentic than human-written brand voice. By contrast, Altay and Gilardi (2024) found that headlines labelled as AI-generated were less likely to be believed and shared, while Toff and Simon (2025) showed that audiences perceived AI-labelled news as less trustworthy even when the underlying articles were not necessarily judged less accurate or less fair. Our findings align more closely with this latter line of work, especially with respect to trust and strategic suspicion, but they also reinforce the idea that disclosure effects are context-dependent rather than universal.

One useful way to interpret this mixed pattern is through research on responses to algorithmic systems. Studies on algorithm aversion suggest that people may avoid algorithmic outputs after perceiving error or inappropriate machine intrusion, whereas research on algorithm appreciation indicates that people sometimes prefer algorithmic over human judgement, particularly when algorithms are associated with consistency or superior performance. This broader literature helps explain why disclosure is not uniformly harmful across contexts: in some settings, AI may cue competence, efficiency, or neutrality, while in others it may cue distance, artificiality, or strategic manipulation. In the present design, which used short persuasive-looking consumer texts, the suspicion route appears to have been more salient than the competence route.

The findings also reinforce the broader literature on mediated authenticity. Authenticity was not determined solely by message content. Instead, it was shaped by contextual framing, expectations about authorship, and perceived communicative intent. Luoma-Aho et al. (2019) show that the same sponsored content can produce different authenticity judgements depending on the valence of the introductory cue, and the present study extends that logic to short written texts and AI disclosure. Here, the AI label functioned as a contextual frame: it did not change the semantic content of the texts, but it changed how that content was interpreted.

The style manipulation was successful, but its evaluative effects were modest. Social-media-style texts produced slightly higher trust and persuasive-intent ratings, while authenticity and ethical concern did not differ reliably. Conversational language may therefore provide relational immediacy while also making strategic intent more noticeable, which helps explain the weaker and more ambivalent style effects.

The results also contribute to discussions of ethical concern. In the PKM tradition, persuasion knowledge includes beliefs not only about tactics and motives but also about the appropriateness of persuasive attempts. The present findings suggest that AI disclosure can increase ethical unease, presumably because some participants interpret AI-generated persuasive-sounding text as less appropriate or less fair. That effect was meaningful, though more moderate than in the earlier, stronger version of the dataset. A recipient may perceive a message as strategic without necessarily judging it unethical. The positive association between persuasive intent and ethical concern supports precisely this distinction.

From a substantive perspective, the study suggests that AI disclosure may operate less as a neutral transparency device and more as an interpretive trigger. This is not necessarily an argument against disclosure. Rather, it implies that disclosure has communicative consequences beyond mere factual clarification. Once readers are told that a text is AI-generated, they may become more attentive to possible motives, more likely to notice persuasive tactics, and more cautious in attributing genuineness to the message. In this sense, transparency does not simply reveal authorship; it may actively reorganise message reception. This observation is consistent with what Toff and Simon (2025) describe as a disclosure dilemma: transparency may be normatively desirable while still reducing trust under some conditions.

The study also has practical implications. Communicators should expect explicit AI disclosure to affect reception, particularly when texts rely on personal voice or sincerity. Clear, standardized marking of AI-generated or materially AI-assisted persuasive outputs could improve transparency, but labels should identify the relevant output and the degree of AI involvement rather than merely attach a vague warning. Such marking should be tested for comprehension because transparency may simultaneously increase informed awareness and reduce trust.

The demographic profile of the sample also matters for interpretation. Because the respondents were predominantly young adults and largely student-based, the present findings may reflect the expectations and interpretive habits of a comparatively digitally literate population. It is plausible that

older samples, professionally different populations, or groups with lower exposure to AI tools would respond differently to authorship disclosure. Accordingly, the present demographic structure should be treated as informative for contextualising the findings, but not as a basis for broad population-level generalisation.

Finally, the present pilot highlights a broader theoretical point. AI disclosure does not simply reveal who produced a text; it changes what kind of social act the text appears to be. In some contexts, that may activate heuristics of objectivity, efficiency, or competence. In others, as in the present case, it may activate persuasion knowledge, lower authenticity, and increase ethical unease. The challenge for future research is therefore not simply to ask whether disclosure is good or bad, but to specify when, for whom, and under what stylistic and contextual conditions disclosure shifts communication towards trust, scepticism, or strategic suspicion.

7. Limitations

The study is small, exploratory and based on single-item measures. It does not provide full construct validation or represent all forms of AI-generated writing. The low-stakes texts help isolate the disclosure cue but limit ecological range, and the design tests explicit disclosure rather than suspected AI authorship. Future studies should use validated multi-item scales, larger samples, mixed-effects models and, where theoretically appropriate, MANOVA or descriptive discriminant analysis to examine the joint pattern of outcomes.

A further limitation concerns the demographic profile. Although age, gender and profession or field of study are included in the present pilot dataset for descriptive purposes, these fields were appended post hoc and were not used as inferential moderators in the reported analyses. Future research should collect and analyse demographic variables prospectively and should explicitly test whether the effects of AI disclosure vary by age, gender, educational background, occupation or AI familiarity.

8. Conclusion

The spread of generative AI has intensified the importance of authorship cues. In this pilot, AI disclosure increased perceived persuasive intent and ethical concern while reducing authenticity and trust. These findings support further research and also point to a policy need: legislation and professional standards should develop faster, clearer rules for identifying targeted AI-generated outputs, especially persuasive or public-facing content. Regulation should specify when disclosure is required, what degree of AI involvement must be stated and how labels can remain understandable without implying that all AI-assisted communication is deceptive. The present study offers a compact experimental basis for testing such disclosure designs.

Acknowledgement

The work on this paper was supported by the Czech Science Foundation (GACR) grant 24-11697S.

References

Altay, S., & Gilardi, F. (2024). People are skeptical of headlines labeled as AI-generated. PNAS Nexus, 3(10), 403. http://doi:10.1093/pnasnexus/pgae403

Campbell, M. C., & Kirmani, A. (2000). Consumers' use of persuasion knowledge: The effects of accessibility and cognitive capacity on perceptions of an influence agent. Journal of Consumer Research, 27(1), 69–83. http://doi:10.1086/314309

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. Journal of Experimental Psychology: General, 144(1), 114–126. http://doi:10.1037/xge0000033

Friestad, M., & Wright, P. (1994). The Persuasion Knowledge Model: How people cope with persuasion attempts. Journal of Consumer Research, 21(1), 1–31. http://doi:10.1086/209380

Kirmani, A., & Campbell, M. C. (2008). I know what you're doing and why you're doing it: The use of the Persuasion Knowledge Model in consumer research. In C. P. Haugtvedt, P. Herr, & F. Kardes (Eds.), Handbook of Consumer Psychology (pp. 549–573). Lawrence Erlbaum.

Kirkby, A., Baumgarth, C., & Henseler, J. (2023). To disclose or not disclose is no longer the question: Effect of AI-disclosed brand voice on brand authenticity and attitude. Journal of Product and Brand Management, 32(7), 1108–1122. http://doi:10.1108/JPBM-02-2022-3864

Logg, J. M., Minson, J. A., & Moore, D. A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. Organizational Behavior and Human Decision Processes, 151, 90–103. http://doi:10.1016/j.obhdp.2018.12.005

Luoma-Aho, V., Pirttimäki, T., Maity, D., Munnukka, J., & Reinikainen, H. (2019). Primed authenticity: How priming impacts authenticity perception of social media influencers. International Journal of Strategic Communication, 13(4), 352–365. http://doi:10.1080/1553118X.2019.1617716

Toff, B., & Simon, F. M. (2025). "Or they could just not use it?": The dilemma of AI disclosure for audience trust in news. The International Journal of Press/Politics, 30(4), 881–903. http://doi:10.1177/19401612251346086

SOCIAL MEDIA BEHAVIOUR AS TRAINING DATA FOR ARTIFICIAL INTELLIGENCE: ETHICAL CHALLENGES AND RISKS

Ivan Katrenčik, Monika Zatrochová

Slovak University of Technology in Bratislava
Institute of Management
ivan.katrencik@stuba.sk, monika.zatrochova@stuba.sk

DOI: 10.35011/IDIMT-2026-81

Keywords

Artificial Intelligence; social media behaviour; ethical AI; AI content generation.

Abstract

Artificial intelligence is no longer only a background technology of digital platforms. It increasingly participates in the production, organisation and circulation of online content. This raises specific ethical questions in social media environments, where user interactions generate large volumes of behavioural and communicative data. The paper examines ethical risks connected with AI-generated content and social media user behaviour. It is based on a structured literature review supported by Scite-assisted source verification. The review identifies four main risk areas: AI-generated misinformation, bias and toxicity in training data and outputs, transparency and labelling of synthetic content, and behavioural modelling as a possible source of manipulation. The findings suggest that ethical risks do not arise only from inaccurate AI outputs. They emerge across a wider communication chain that includes data sources, model behaviour, platform incentives and user interpretation. Social media behaviour therefore has a dual role: it can serve as input for AI systems, but it can also become a target of AI-driven influence. The paper concludes that transparency, labelling and moderation are necessary, but insufficient without broader accountability mechanisms involving AI developers, platform operators and regulators. The contribution of the paper lies in proposing a sociotechnical perspective on the ethical evaluation of AI-generated content in digital communication.

1. Introduction

Artificial intelligence has moved from being a background technology of digital platforms to an active producer, organiser and moderator of online communication. This shift is especially visible in social media environments, where AI is used not only to recommend or moderate content, but increasingly also to generate text, images, audio and video. The ethical relevance of this development lies in the character of social media itself. Social media data are not neutral records of human behaviour. They include preferences and opinions, but also emotional reactions, stereotypes, conflicts, misinformation and strategic attempts to influence others. This is consistent with broader discussions on ethical data use, which emphasise that the way data are collected, interpreted and processed has direct implications for responsibility, privacy and trust in digital environments (Lisnik & Kuperová, 2023).

Social media data are not neutral records of human behaviour. They include preferences and opinions, but also emotional reactions, stereotypes, conflicts, misinformation and strategic attempts to influence others.

The broader AI ethics literature identifies transparency, fairness, privacy, responsibility and non-maleficence as key principles of responsible AI governance (Jobin et al., 2019). In social media contexts, however, these principles are difficult to apply in a straightforward way. Platforms are shaped by engagement metrics, commercial incentives and continuous behavioural feedback. AI systems operating in this environment do not merely process information; they participate in selecting, ranking, producing and circulating communication.

Recent studies show that generative AI may intensify existing ethical problems in digital communication. LLMs can contribute to the production and dissemination of misleading content in online environments where social media already facilitates rapid information circulation (Shah et al., 2025). At the same time, generative models may reproduce bias, stereotypes and toxic patterns embedded in large-scale textual data (Babonnaud et al., 2024). A further concern is behavioural modelling. Generative AI can support personalised persuasion by adapting messages to inferred psychological characteristics of users (Matz et al., 2024).

The aim of this paper is to examine ethical risks connected with AI-generated content and social media user behaviour. The paper is based on a structured literature review supported by Scite-assisted source verification. It focuses on four interrelated areas: AI-generated misinformation, bias and toxicity in data and outputs, transparency and labelling of synthetic content, and behavioural modelling as a potential source of manipulation.

2. Research methodology

This paper is based on a structured literature review. This approach was selected because the aim of the study is not to collect primary empirical data, but to synthesise existing knowledge on the ethical risks connected with AI-generated content, behavioural modelling and social media user behaviour. Literature-based approaches are commonly used in AI ethics research to identify dominant principles, conceptual tensions and implementation challenges (Khan et al., 2022).

The literature search was conducted with the support of Scite, which was used both for source identification and citation-context verification. Search terms combined the following concepts: “AI ethics”, “generative AI”, “AI-generated content”, “social media”, “user behaviour”, “misinformation”, “bias”, “transparency”, “labelling”, “human autonomy” and “behavioural modelling”. The search focused mainly on publications from 2019 onwards.

Sources were selected if they addressed at least one of the central themes of the paper and provided a theoretical argument, empirical evidence, review synthesis or governance perspective relevant to the research objective. The selected literature was analysed thematically. Four analytical categories were derived from the review: misinformation and AI-generated content, bias and toxicity in training data and outputs, transparency and labelling of synthetic content, and behavioural modelling as a potential source of manipulation.

3. Results

3.1. AI-generated content and misinformation

The first thematic finding concerns misinformation. The reviewed literature suggests that generative AI changes the conditions under which misleading content can be produced and circulated online. Shah et al. (2025) describe LLMs as double-edged tools in the context of misinformation and disinformation: they may support detection and analysis, but they may also be used to generate or amplify misleading content. This is particularly relevant in social media environments, where information spreads quickly and where the boundary between accurate and false content may become blurred during everyday user interaction.

Radivojevic et al. (2024) similarly argue that LLMs may reshape digital discourse on social media while raising risks related to propaganda, privacy and bias. The ethical risk is therefore not limited to factual error. AI-generated content may appear coherent, fluent and socially plausible, which can make misleading communication harder for users to evaluate. Shin et al. (2024) add that the impact of generative AI misinformation also depends on how users process and interpret AI-generated information. Misinformation should therefore be understood not only as a content-level problem, but also as a user- and platform-level problem.

3.2. Bias and toxicity in training data and generated outputs

The second thematic finding concerns bias and toxicity. LLMs rely on large volumes of textual data, including publicly available internet and social media content. Linegar et al. (2023) note that LLMs learn from sources such as blogs, social media, books and articles, and warn that biases present in such data may be reproduced in model outputs. This is important because social media communication is not a balanced representation of society. It often contains emotionally charged, polarised or discriminatory language.

Babonnaud et al. (2024) show that LLMs may generate toxic, stereotypical and ethically problematic outputs, partly because the textual data used for training include examples of harmful or prejudicial communication. The issue is not limited to offensive wording. Peters and Matz (2024) demonstrate that LLMs can infer psychological dispositions from social media posts, while also reporting systematic differences in estimation accuracy across demographic groups. Bias may therefore affect not only what AI systems generate, but also how they classify, interpret and potentially target users.

3.3. Transparency, labelling and user trust

The third thematic finding concerns transparency and labelling. In social media environments, users may not know whether they are interacting with human-created, AI-assisted or fully AI-generated content. Epstein et al. (2023) argue that even defining what should count as AI-generated content is not straightforward, because there is a continuum between different forms of algorithmic involvement. This makes transparency more complicated than a simple distinction between “human” and “machine-generated” communication.

Labelling may improve user awareness, but the reviewed literature suggests that it remains only a partial solution. Liu et al. (2023) show that AIGC labels can influence how users perceive automated news and may support more reflective information processing. At the same time, Fisher et al. (2024) argue that transparency and labelling should be understood as limited governance tools rather than complete solutions to misinformation or manipulation. Their analysis further highlights the role of

social media platforms, which remain the main distribution environments for synthetic content and therefore carry significant responsibility for its moderation and disclosure.

3.4. Behavioural modelling and manipulation risks

The fourth thematic finding concerns behavioural modelling and manipulation. Social media platforms generate large amounts of behavioural data that may be used to infer preferences, interests and psychological characteristics of users. Peters and Matz (2024) show that LLMs can infer psychological dispositions from social media posts and discuss the risks associated with automated psychological assessment. Their findings suggest that AI systems may not only analyse online behaviour, but also classify and interpret users in ways that may influence later communication or targeting.

The ethical concern becomes stronger when behavioural inference is combined with generative AI. Matz et al. (2024) provide empirical evidence that generative AI can support personalised persuasion at scale by producing messages adapted to psychological profiles. In this context, the boundary between personalisation and manipulation becomes increasingly difficult to define. Prunkl (2022) similarly argues that AI systems may affect human autonomy by shaping information environments and influencing preference formation. This is especially relevant in social media ecosystems, where recommendation systems and engagement-based communication already play a central role in directing user attention and interaction.

3.5. Summary of identified ethical risk areas

The structured literature review identified several interrelated ethical risk areas. These risks do not operate independently; rather, they overlap across the data, model, content, platform and user levels. Table 1 summarises the main findings of the review and indicates their relevance to social media user behaviour.

Table 1. Summary of ethical risks identified in the literature review

Ethical risk area	Main finding from the literature	Relevance to social media user behaviour	Selected supporting sources
AI-generated misinformation	Generative AI can produce fluent, persuasive and potentially misleading content. The risk is not only factual error, but also the scale and credibility of generated communication.	Social media enables rapid sharing and engagement-based visibility, which may increase the risk that misleading AI-generated content circulates widely.	Radivojevic et al. (2024); Shah et al. (2025); Monteith et al. (2023); Shin et al. (2024)
Bias and toxicity	LLMs may reproduce biases, stereotypes and toxic patterns present in large-scale internet and social media data. Bias can appear even when prompts are not explicitly harmful.	Social media communication contains emotionally charged, polarised and sometimes discriminatory content that can become part of the data environment from which AI systems learn.	Linegar et al. (2023); Babonnaud et al. (2024); Chen et al. (2023); Peters & Matz (2024)

Ethical risk area	Main finding from the literature	Relevance to social media user behaviour	Selected supporting sources
Transparency and labelling	Labelling AI-generated content may improve user awareness, but its effectiveness depends on wording, design, placement and user interpretation.	Users may not know whether content is human-created, AI-assisted or fully synthetic, which can affect credibility judgments and trust.	Wittenberg et al. (2024); Epstein et al. (2023); Liu et al. (2023); Gamage et al. (2025)
Behavioural modelling and manipulation	AI systems can infer user characteristics and generate psychologically tailored persuasive messages. This increases the scalability of personalised persuasion.	Social media produces behavioural traces that can be used to predict preferences, personalise content and potentially influence attitudes or decisions.	Matz et al. (2024); Peters & Matz (2024); Prunkl (2022); Montag et al. (2025)
Governance and accountability	General ethical principles are insufficient unless they are connected to documentation, moderation, platform responsibility and mechanisms for inspection or modification.	Responsibility is distributed across AI developers, platforms and regulators, especially when AI-generated content circulates through social media infrastructures.	Ferrari et al. (2023); Fisher et al. (2024); Wahle et al. (2023)

Source: authors own research

4. Discussion

The synthesis presented in Table 1 shows that the ethical risks of AI-generated content in social media environments are interconnected rather than isolated. Misinformation, bias, transparency and behavioural modelling form a chain in which social media behaviour may become both an input for AI systems and a target of AI-driven influence. A misleading AI-generated post, for example, may be problematic not only because it contains false information, but also because it is fluently written, may reproduce biased patterns from training data, circulates through engagement-based platform mechanisms and can be adapted to the characteristics of particular users.

This interpretation is consistent with recent literature on generative AI and misinformation. Shah et al. (2025) describe LLMs as "double-edged" tools in the context of misinformation and disinformation, since they may support both detection and generation of misleading content. Radivojevic et al. (2024) similarly argue that LLMs can reshape digital discourse on social media while raising risks related to propaganda, privacy and bias. Our review extends this discussion by showing that misinformation should not be analysed separately from platform dynamics and user behaviour. In social media environments, the credibility and impact of content depend not only on its factual accuracy, but also on how it is presented, shared and interpreted.

The findings also correspond with studies on bias and stereotypes in large language models. Babonnaud et al. (2024) show that concerns about LLMs include toxic, stereotypical and ethically problematic outputs, while Linegar et al. (2023) emphasise that LLMs depend on large-scale textual data and may reproduce biases present in those data. In this paper, these findings are interpreted through the specific lens of social media behaviour. Social media data are not only large; they are also shaped by emotion, conflict, selective visibility and platform-specific communication norms. This makes them ethically sensitive as training or optimisation data.

The results further support the argument that transparency and labelling are necessary but limited governance tools. Epstein et al. (2023) show that even the terminology used for labelling AI-generated content is not straightforward, because there is a continuum between human-created, AI-assisted and fully AI-generated communication. Liu et al. (2023) suggest that AIGC labelling can affect users' perception of automated news, but Fisher et al. (2024) caution that labelling remains only a partial solution. Our findings are aligned with this view. Labelling may improve user awareness, but it does not automatically solve problems of bias, misinformation or manipulation.

The issue of transparency is also connected with authorship and integrity. In academic and professional contexts, the use of AI-generated outputs raises questions about who should be considered responsible for the final content and how AI assistance should be disclosed. Lisnik et al. (2024) discuss this problem in relation to authorial integrity in contemporary academic practice. Although the context of their study is academic work, the underlying ethical tension is similar in digital communication: users and institutions need clearer rules for distinguishing between human authorship, AI assistance and automated content production.

The most sensitive implication concerns behavioural modelling. Matz et al. (2024) provide empirical evidence that generative AI can support personalised persuasion at scale, while Peters and Matz (2024) show that LLMs can infer psychological dispositions from social media posts. These findings are especially relevant to the topic of this paper because they indicate a shift from content generation to influence optimisation. If AI systems are able to infer user characteristics and generate tailored persuasive messages, the ethical boundary between personalisation and manipulation becomes increasingly difficult to define. This is particularly important in social media environments, where behavioural data are continuously produced, and platform incentives often reward engagement.

From a governance perspective, the findings support the need for more concrete mechanisms of accountability. Ferrari et al. (2023) argue that generative AI governance requires conditions under which systems can be observed, inspected and modified. Wahle et al. (2023) similarly emphasise the importance of transparent reporting of AI-generated content. These perspectives suggest that general ethical principles, such as fairness, transparency and responsibility, are not sufficient unless they are connected to documentation, moderation practices, disclosure rules and institutional oversight. In the context of social media, responsibility should therefore be distributed across model developers, platform operators and regulators, rather than transferred mainly to users.

The study has several limitations. First, it is based on a structured literature review and does not include primary empirical data. Therefore, it cannot measure how users actually respond to AI-generated content in a specific platform or national context. Second, the literature on generative AI is developing very quickly, and some relevant studies may become outdated as technologies, platform policies and regulations change. Third, several sources included in the review are conceptual papers, preprints or early-stage empirical studies, which reflects the emerging nature of the field. Fourth, the review focuses mainly on English-language academic literature, which may limit its ability to capture regional, cultural or linguistic differences in social media behaviour and AI governance.

Future research should address these limitations in several ways. Empirical studies could examine how users perceive labelled and unlabelled AI-generated content across different social media platforms. Comparative research could investigate whether users from different age groups, countries or educational backgrounds differ in their ability to recognise synthetic content or resist personalised persuasion. Another important direction is the study of platform-level governance: how social media companies implement AI content labelling, moderation and transparency obligations in practice. Finally, future research should examine the interaction between psychological profiling, generative AI and persuasive communication, especially in politically or commercially sensitive contexts.

Overall, the discussion shows that ethical risks in AI-generated social media content cannot be solved by focusing on one element alone. Improving model safety, labelling synthetic content, educating users and regulating platforms are all necessary, but none of them is sufficient in isolation. A sociotechnical approach is required, because the risks emerge from the interaction between data, models, platforms and users. The central ethical question is therefore not only whether AI-generated content is accurate, but how it is produced, disclosed, distributed and used to influence human behaviour.

5. Conclusion

This paper examined ethical risks related to AI-generated content and social media user behaviour through a structured literature review. The findings show that the main ethical challenges are not limited to the accuracy of generated outputs. They also concern the origin of training data, the reproduction of bias and toxicity, the transparency of AI-generated communication, and the possible use of behavioural modelling for persuasive or manipulative purposes.

The review suggests that social media behaviour has a dual role in AI systems. It can serve as a data source from which AI models learn, but it can also become a target of AI-driven influence. This makes the ethical evaluation of AI-generated content more complex, because risks emerge across the whole communication chain: data, model, content, platform and user.

A key conclusion is that transparency, labelling and moderation are necessary, but not sufficient on their own. Responsible use of AI in social media environments requires stronger accountability from AI developers, platform operators and regulators. Future research should focus especially on user responses to labelled AI-generated content, platform governance practices, and the ethical boundaries of personalised AI-driven persuasion.

Acknowledgement

The contribution is a partial output of the KEGA research task no. KEGA č. 031STU-4/2025 conducted at the Institute of Management of the STU in Bratislava.

References

Babonnaud, W., Delouche, E., & Lahlouh, M. (2024). The bias that lies beneath: Qualitative uncovering of stereotypes in large language models. In Linköping Electronic Conference Proceedings, 208, 195–203. https://doi.org/10.3384/ecp208022

Chen, C., Fu, J., & Lyu, L. (2023). A pathway towards responsible AI generated content. *arXiv*. https://doi.org/10.48550/arXiv.2303.01325

Epstein, Z., Fang, M. C., & Arechar, A. A. (2023). What label should be applied to content produced by generative AI? OSF Preprints. https://doi.org/10.31234/osf.io/v4mfz

Ferrari, F., van Dijck, J., & van den Bosch, A. (2023). Observe, inspect, modify: Three conditions for generative AI governance. New Media & Society, 27(5), 2788–2806. https://doi.org/10.1177/14614448231214811

Fisher, S. A., Howard, J. W., & Kira, B. (2024). Moderating synthetic content: The challenge of generative AI. Philosophy & Technology, 37(4). https://doi.org/10.1007/s13347-024-00818-9

Gamage, D., Sewwandi, D., & Zhang, M. (2025). Labeling synthetic content: User perceptions of label designs for AI-generated content on social media. In Proceedings of the ACM Conference on Human Factors in Computing Systems. https://doi.org/10.1145/3706598.3713171

Khan, A. A., Badshah, S., & Liang, P. (2022). Ethics of AI: A systematic literature review of principles and challenges. In Proceedings of the 26th International Conference on Evaluation and Assessment in Software Engineering (pp. 383–392). ACM. https://doi.org/10.1145/3530019.3531329

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. Nature Machine Intelligence, 1(9), 389–399. https://doi.org/10.1038/s42256-019-0088-2

Lechterman, T. M. (2022). The concept of accountability in AI ethics and governance. In The Oxford Handbook of AI Governance (pp. 164–182). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780197579329.013.10

Linegar, M., Kocielnik, R., & Alvarez, R. M. (2023). Large language models and political science. Frontiers in Political Science, 5. https://doi.org/10.3389/fpos.2023.1257092

Lisnik, A., & Kuperová, M. (2023). Ethical aspects of working with data. In IDIMT-2023: New challenges for ICT and management (pp. 357–365). Trauner Verlag. https://doi.org/10.35011/IDIMT-2023-357

Lisnik, A., Majerník, M., & Salanci, J. (2024). Ethical perspectives and the issue of authorial integrity in the use of AI in contemporary academic practice. In IDIMT-2024: New challenges for ICT and management. Trauner Verlag. https://doi.org/10.35011/IDIMT-2024-405

Liu, Y., Wang, S., & Yu, G. (2023). The nudging effect of AIGC labeling on users' perceptions of automated news: Evidence from EEG. Frontiers in Psychology, 14. https://doi.org/10.3389/fpsyg.2023.1277829

Matz, S., Teeny, J. D., & Vaid, S. S. (2024). The potential of generative AI for personalized persuasion at scale. Scientific Reports, 14(1). https://doi.org/10.1038/s41598-024-53755-0

Montag, C., Yang, H., & Wu, A. M. S. (2025). The role of artificial intelligence in general, and large language models specifically, for understanding addictive behaviors. Annals of the New York Academy of Sciences, 1548(1), 5–11. https://doi.org/10.1111/nyas.15337

Monteith, S., Glenn, T., & Geddes, J. (2023). Artificial intelligence and increasing misinformation. The British Journal of Psychiatry, 224(2), 33–35. https://doi.org/10.1192/bjp.2023.136

Peters, H., & Matz, S. (2024). Large language models can infer psychological dispositions of social media users. PNAS Nexus, 3(6). https://doi.org/10.1093/pnasnexus/pgae231

Prunkl, C. (2022). Human autonomy in the age of artificial intelligence. Nature Machine Intelligence, *4*(2), 99–101. https://doi.org/10.1038/s42256-022-00449-9

Radivojevic, K., Clark, N. J., & Brenner, P. (2024). LLMs among us: Generative AI participating in digital discourse. AAAI Spring Symposium Series, 3(1), 209–218. https://doi.org/10.1609/aaaiss.v3i1.31202

Shah, S. B., Thapa, S., & Acharya, A. (2025). Navigating the web of disinformation and misinformation: Large language models as double-edged swords. IEEE Access, 13, 169262–169282. https://doi.org/10.1109/ACCESS.2024.3406644

Shin, D., Koerber, A., & Lim, J. S. (2024). Impact of misinformation from generative AI on user information processing: How people understand misinformation from generative AI. New Media & Society, 27(7), 4017–4047. https://doi.org/10.1177/14614448241234040

Wahle, J. P., Ruas, T., & Mohammad, S. M. (2023). AI usage cards: Responsibly reporting AI-generated content. Authorea. https://doi.org/10.22541/au.168323192.20543041/v1

Wittenberg, C., Epstein, Z., & Berinsky, A. J. (2024). Labeling AI-generated content: Promises, perils, and future directions. MIT Generative AI. https://doi.org/10.21428/e4baedd9.0319e3a6

Note: During the preparation of this paper, the authors used ChatGPT as an auxiliary tool for language refinement, structuringof arguments and support in the formulation of selected sections. Scite was used to support literature search and citation-context verification. The authors remain fully responsible for the final content, interpretation of sources and all conclusions presented in the paper.

DATA, AI AND DIGITAL-DRIVEN TRANSFORMATION: SHAPING SUSTAINABLE DIGITAL FUTURES

DYNAMIC DIGITAL CAPABILITIES FOR DIGITAL RESILIENCE OF SMES: CONCEPTUAL FOUNDATIONS AND RESEARCH DESIGN

Andreja Pucihar, Mirjana Kljajic Borstnar

University of Maribor
Faculty of organizational sciences
andreja.pucihar@um.si, mirjana.kljajic@um.si

DOI: 10.35011/IDIMT-2026-91

Keywords

Resilience; digital transformation; dynamic capabilities; SME; multi-criteria decision modelling.

Abstract

Small and medium-sized enterprises (SMEs) operate in increasingly turbulent digital environments shaped by rapid technological change, growing data dependencies, cybersecurity threats, and expanding regulatory requirements. To remain competitive, SMEs must develop digital resilience—the ability to anticipate, absorb, adapt to, and transform in response to disruptions and continuous change. However, existing digital maturity and resilience frameworks provide limited support for SMEs, as they often focus on large organizations, emphasize diagnostic measurement, and insufficiently address ecosystem and regulatory influences. This paper develops a conceptual framework for understanding digital resilience in SMEs by integrating dynamic capabilities theory, organizational resilience, and digital sovereignty. The framework identifies four determinant dimensions - organizational, technological, relational and ecosystem, and institutional and regulatory - that influence the development of dynamic digital capabilities and corresponding resilience capacities. Building on this conceptualization, the paper outlines design principles for an explainable assessment approach based on qualitative multi-criteria decision modelling (DEX). The paper contributes by advancing an integrated view of SME digital resilience and by providing a foundation for the future development of assessment and decision-support tools that can help SMEs strengthen resilience in digitally intensive and regulation-driven environments.

1. Introduction

SMEs form the backbone of European economies, accounting for over 99% of all enterprises in the EU. At the same time, they face increasing digital, environmental, and regulatory turbulence (OECD, 2023). The rapid diffusion of AI, growing data dependencies, cybersecurity threats, platform concentration, and expanding EU regulatory frameworks—including the AI Act (AI Act, 2024), CSRD (CSRD Directive, 2022), the Data Act (Data Act, 2023), and the Data Governance Act (Data Governance Act, 2022) - create new challenges for SME adaptability and competitiveness. Although some of these regulations primarily target larger organizations, SMEs are increasingly affected through technology use, data governance requirements, and supply-chain compliance obligations.

While digital transformation creates opportunities for innovation and competitiveness (Marolt et al., 2024; Vial, 2019), it also increases technological dependency, regulatory exposure, and vulnerability to disruptions in cloud infrastructures, data flows, and AI ecosystems (Marolt et al., 2024; Merlo et al., 2025). Consequently, digital sovereignty - the ability to retain strategic autonomy over data, technologies, and digital infrastructures - has become an important concern for SMEs (Falkner et al., 2024; OECD, 2023).

Organizational resilience refers to the ability to anticipate, adapt to, and sustain performance under continuous change (Burnard & Bhamra, 2011). Digital resilience further highlights the dual role of digital technologies as both sources of vulnerability and enablers of adaptation (Warner & Wäger, 2019). These challenges are particularly pronounced in SMEs due to limited resources and dependence on external digital infrastructures and ecosystem actors.

Developing resilience requires SMEs to understand their capabilities and identify priorities for improvement. However, existing digital maturity models remain predominantly diagnostic, quantitatively oriented, and designed for larger enterprise contexts. They provide limited support for capability development and strategic prioritization in SME environments (Kljajić Borštnar & Pucihar, 2021; Mithas et al., 2013; Schumacher et al., 2016). Moreover, organizational, technological, ecosystem, and regulatory factors are often examined separately, while emerging concerns such as digital sovereignty, data governance, and regulatory compliance remain insufficiently integrated (Marolt et al., 2024). As a result, there is a need for integrated, explainable, and actionable frameworks that support the assessment and development of SME digital resilience.

To address this gap, this paper develops a conceptual framework for understanding and assessing digital resilience in SMEs. Drawing on organizational resilience theory, dynamic capabilities theory, and digital sovereignty, it identifies determinant dimensions that shape the development of dynamic digital capabilities and corresponding resilience capacities. Building on this foundation, it derives design principles for a digital resilience assessment approach tailored to SME needs.

The aim of this paper is therefore twofold:

1) to conceptualize digital resilience in SMEs as a capability-driven, multi-dimensional construct linking determinants, dynamic capabilities, and resilience outcomes; and
2) to derive the design principles for the future development of an explainable assessment approach that supports SME strategic decision-making.

The remainder of the paper is organized as follows. Section 2 reviews the theoretical foundations. Section 3 presents the research design and methodological approach. Section 4 introduces the conceptual framework. Section 5 presents the discussion and conclusion.

2. Theoretical background

2.1. Dynamic capabilities and digital resilience: SMEs perspective

The dynamic capabilities framework (Teece, 2007) argues that competitive advantage in turbulent environments depends on an organization's ability to sense changes, seize opportunities, and reconfigure its resource base. In digitally intensive contexts, these capabilities extend to the adaptation of digital assets, processes, and knowledge, often conceptualized as dynamic digital capabilities (Warner & Wäger, 2019).

From an SME perspective, the development of dynamic capabilities is shaped by resource constraints, limited formalization, and dependence on external partners and digital infrastructures (Marolt et al.,

2024). As a result, SMEs rely on flexible organizational practices, continuous learning, and ecosystem relationships to respond to digital transformation pressures (Gonzalez-Varona et al., 2021; Saeedikiya et al., 2024; Sawang & Sornlertlamvanich, 2026).

This perspective is closely linked to organizational resilience, conceptualized as a meta-capability comprising anticipatory, absorptive, adaptive, and transformative capacities. In digitally turbulent environments, resilience is shaped by the dual role of digital technologies: they introduce vulnerabilities such as cybersecurity risks, platform dependencies, and data exposure while enabling monitoring, redundancy, and reconfiguration (Warner & Wäger, 2019).

Dynamic capabilities theory explains how such resilience capacities develop. Through sensing, seizing, and transforming, firms reconfigure resources and competences in response to technological, regulatory, and environmental change (Teece, 2007; Ellström et al., 2022). Digital resilience can therefore be understood as an outcome of dynamic capability development, enabling organizations to anticipate risks, absorb shocks, adapt operations, and transform business models (Sagala & Őri, 2026).

A further concept relevant to this discussion is digital sovereignty, understood as the capacity of an organization to retain meaningful control over its digital assets, data, and infrastructural dependencies (Ko et al., 2023; OECD, 2023). For SMEs, participation in digital ecosystems often entails dependence on platform providers, cloud services, and external intermediaries. Digital sovereignty therefore does not imply isolation, but the ability to maintain capability-based autonomy within interconnected digital environments.

Together, these perspectives suggest a strong linkage between dynamic capabilities, organizational resilience, and digital sovereignty. This paper builds on their integration and conceptualizes SME digital resilience as the outcome of dynamic digital capabilities shaped by organizational, technological, relational, and institutional conditions.

2.2. Digital Maturity Models: State of the art and gaps

Digital maturity models (DMMs) provide structured frameworks for assessing an organization's digital development and identifying improvement pathways. Examples include the Digital Maturity Model (Gill & VanBoskirk, 2016), the IMPULS Industry 4.0 Maturity Index (Schumacher et al., 2016), SME-oriented assessment models (Kljajić Borštnar & Pucihar, 2021), and the Open Digital Maturity Assessment tool (Open DMAT, n.d.). Despite differences in scope, these models typically combine assessment dimensions, maturity levels, and indicators used to evaluate organizational characteristics.

Despite their usefulness, existing DMMs show important limitations from the perspective of SME digital resilience. Many rely on quantitative or semi-quantitative indicators that are difficult to apply where formalized data, processes, and measurement systems are limited. Moreover, most frameworks are primarily diagnostic: they identify current maturity levels but provide limited support for capability development and strategic prioritization. Ecosystem and regulatory dimensions - such as platform dependence, supply chain exposure, compliance requirements, and broader institutional conditions - also remain underrepresented (Mithas et al., 2013).

These limitations are particularly relevant as SMEs make strategic decisions under technological uncertainty and regulatory change. To support strategic action, maturity models should help organizations not only classify current states but also interpret development needs and identify pathways for capability building (Schumacher et al., 2016). This highlights the need for assessment approaches that are more qualitative, explainable, and context-sensitive than many existing models.

2.3. The EU regulatory context and SME exposure

A growing number of EU initiatives shape how organizations adopt, govern, and use digital technologies. The AI Act (2024), the Corporate Sustainability Reporting Directive (CSRD) (CSRD Directive, 2022), the Data Governance Act (2022), and the Data Act (2023) introduce new requirements related to AI deployment, sustainability reporting, data sharing, and data governance. These developments interact with ecosystem-level dynamics such as platform concentration, supply chain digitalization, and cybersecurity threats, generating pressures that existing maturity models do not systematically address.

In this context, regulatory awareness and compliance capacity should be viewed not merely as administrative requirements but as organizational capabilities that shape how SMEs respond to digital transformation. The institutional and regulatory environment therefore represents an important dimension of digital resilience alongside organizational, technological, and ecosystem-related determinants. For SMEs, maintaining competitiveness, ensuring compliance, and preserving strategic autonomy become closely interrelated concerns.

3. Research framework and methodology

3.1. Research approach: Design Science Research

This study follows a design science research (DSR) approach. This approach is particularly suitable for addressing complex, practice-oriented problems through the development of various types of IT artefacts (Hevner et al., 2004). In this context, the artefact under development is a conceptual and methodological foundation for a digital resilience assessment approach tailored to SMEs.

DSR emphasizes the iterative interplay between problem understanding, artefact design, and evaluation within relevance and rigor cycles. It has been widely applied in the development of maturity models and decision-support tools, where the objective is not only to explain phenomena but also to provide structured and usable solutions to real-world problems (Kljajić Borštnar & Pucihar, 2021).

In line with this paradigm, the present study focuses on the early design stage of the artefact. Rather than presenting a fully operational model, it develops conceptual structure, identifies key design requirements, and outlines the methodological approach that will guide subsequent development and validation phases.

3.2. Multi-Criteria Decision Modelling as the artefact design approach

Within the DSR framework, the future operationalization of the assessment artefact is grounded in multi-criteria decision modelling (MCDM). The assessment of digital resilience in SMEs represents a multi-dimensional problem involving organizational, technological, ecosystem, and regulatory factors that cannot be meaningfully reduced to a single criterion. MCDM therefore provides a suitable structure for organizing and evaluating these interrelated dimensions.

Among MCDM approaches, this study adopts the DEX (Decision EXpert) methodology as the principal design method for the future assessment model (Bohanec et al., 2013)). DEX uses hierarchical decomposition of decision problems and rule-based aggregation of attributes, enabling transparent and explainable evaluation. Its qualitative value scales are particularly suitable for SME contexts, where relevant information is often difficult to quantify.

At the current stage, DEX serves primarily as a guiding design logic for structuring the assessment framework. The detailed specification of attributes, value scales, and decision rules remains part of future research.

4. Conceptual framework for SME digital resilience

4.1. Overview of the framework

Building on the theoretical foundations and design considerations outlined in the previous sections, this section presents a preliminary conceptual framework for understanding digital resilience in SMEs. The framework represents the central outcome of the design-oriented analysis and provides a structured view of digital resilience as a multi-dimensional and capability-driven phenomenon.

The framework integrates insights from dynamic capabilities theory (Duchek, 2020; Hillmann & Guenther, 2021; Teece, 2007; Warner & Wäger, 2019) and digital sovereignty (Falkner et al., 2024), and conceptualizes digital resilience as an emergent outcome of interactions between organizational conditions, capability development processes, and resilience capacities. It is structured around three interconnected layers: (1) determinant dimensions shaping organizational conditions, (2) dynamic digital capabilities as enabling mechanisms, and (3) resilience capacities as observable outcomes.

4.2. Determinant dimensions of digital resilience

The identification of determinant dimensions is grounded in a synthesis of prior research on digital transformation, SME digital maturity, and organizational resilience, as discussed in Section 2. Existing studies consistently highlight organizational, technological, ecosystem, and institutional factors as key influences on digital capability development in SMEs (Kljajić Borštnar & Pucihar, 2021; Marolt et al., 2024).

Building on these insights and guided by the integration of dynamic capabilities and resilience perspectives, this paper proposes four broad determinant dimensions as an initial conceptual structuring of SME digital resilience. These dimensions should be understood as analytically distinct but interrelated, reflecting the systemic and multi-level nature of SME digital transformation:

- *Organizational determinants*, including leadership commitment, digital culture, skills, and learning capacity, which shape the internal readiness of SMEs to engage with digital and regulatory change.
- *Technological determinants*, encompassing IT infrastructure, data management, cybersecurity, and AI readiness, which influence the technical capacity for resilient and adaptive operations.
- *Relational and ecosystem determinants*, covering partnerships, supply chain integration, and platform dependencies, and sector-specific ecosystem requirements reflecting the embeddedness of SMEs within broader digital ecosystems.
- *Institutional and regulatory determinants*, including compliance capacity, regulatory awareness, and access to support mechanisms, capturing the influence of the external regulatory and policy environment.

Together, these dimensions provide a structured lens for analyzing how SMEs develop the conditions necessary for building digital resilience.

4.3. Linking determinants to dynamic digital capabilities

The framework proposes that the identified determinant dimensions influence the development of dynamic digital capabilities, which act as the central mechanism through which SMEs respond to digital and regulatory turbulence.

In line with dynamic capabilities theory, these capabilities enable SMEs to sense emerging technological, regulatory, and market developments; seize opportunities and respond to risks; reconfigure organizational and technological resources.

Each determinant dimension contributes to this process in a different way. Organizational determinants support learning and responsiveness, technological determinants enable absorptive and operational capacity, ecosystem determinants shape adaptability through inter-organizational relationships, and institutional determinants influence the ability to anticipate and respond to regulatory requirements. Through these interactions, dynamic digital capabilities act as the micro-foundations that translate structural conditions into adaptive organizational responses.

4.4. Resilience capacities as outcomes

At the outcome level, dynamic capabilities translate into resilience capacities, reflecting not only how SMEs are able to respond to disruption and change, but also the extent to which they can sustain and scale those responses. Drawing on organizational resilience literature, four interrelated capacities are considered: anticipatory capacity, referring to the ability to identify and interpret emerging risks and opportunities; absorptive capacity, capturing the ability to withstand shocks without significant performance loss; adaptive capacity, enabling incremental adjustments to changing conditions; transformative capacity, allowing more fundamental reconfiguration of business models and digital architecture.

Digital resilience is thus conceptualized not as a static maturity level, but as an evolving outcome of capability development shaped by multiple interacting determinants.

4.5. Conceptual implications for assessment design

The proposed framework provides the conceptual basis for developing a structured digital resilience assessment model. By linking determinants, capabilities, and resilience outcomes, it offers a coherent structure that can be translated into an evaluative model using multi-criteria decision modelling techniques.

In particular, the framework suggests that effective assessment approaches should: capture multiple interrelated dimensions rather than single indicators; reflect capability development processes rather than static maturity states; incorporate qualitative and context-sensitive evaluation and support interpretation and strategic decision-making. At this stage, these implications remain conceptual and serve as guidance for the subsequent operationalization of the assessment approach.

5. Discussion and conclusion

The proposed framework contributes to literature by integrating dynamic capabilities, organizational resilience, and digital sovereignty into a coherent conceptualization of SME digital resilience. It positions digital resilience as an evolving outcome of capability development and highlights the role of organizational, technological, ecosystem, and institutional determinants in shaping resilience

capacities. By explicitly incorporating regulatory exposure and external dependencies, the framework extends existing resilience perspectives that focus primarily on internal organizational factors.

From a practical perspective, the framework provides a structured basis for the future development of digital resilience assessment approaches for SMEs. By linking determinants, dynamic capabilities, and resilience outcomes, it supports a more comprehensive understanding of resilience than traditional maturity-oriented assessments and offers a foundation for explainable decision-support tools.

This study is conceptual and has not yet been empirically validated. Future research should therefore focus on validating the proposed framework, refining determinant structures and relationships, and developing an operational assessment model, including the specification of attributes, value scales, and decision rules. Further work should also evaluate the usability and decision-support value of such approaches in SME contexts.

Overall, the paper establishes a conceptual and methodological foundation for understanding and assessing digital resilience in SMEs operating in increasingly complex digital and regulatory environments. Rather than presenting a completed assessment model, it provides a basis for future empirical validation and operationalization.

Acknowledgement

This research was supported by the Slovenian Research and Innovation Agency (ARIS): Program No. P5-0018; and Slovenian Research and Innovation Agency (ARIS) and Ministry of Digital Transformation of Republic of Slovenia: Grant No. V2-24066.

References

AI Act. (2024). Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act) (Text with EEA relevance). https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng.

Bohanec, M., Žnidaršič, M., Rajkovič, V., Bratko, I., & Zupan, B. (2013). DEX Methodology: Three Decades of Qualitative Multi-Attribute Modeling. Informatica, 37(1), 49–54.

Burnard, K., & Bhamra, R. (2011). Organisational resilience: Development of a conceptual framework for organisational responses. International Journal of Production Research, 49(18), 5581–5599. https://doi.org/10.1080/00207543.2011.563827

CSRD Directive. (2022). Directive (EU) 2022/2464 of the European Parliament and of the Council of 14 December 2022 amending Regulation (EU) No 537/2014, Directive 2004/109/EC, Directive 2006/43/EC and Directive 2013/34/EU, as regards corporate sustainability reporting (Text with EEA relevance). http://data.europa.eu/eli/dir/2022/2464/oj

Data Act. (2023). Regulation (EU) 2023/2854 of the European Parliament and of the Council of 13 December 2023 on harmonised rules on fair access to and use of data and amending Regulation (EU) 2017/2394 and Directive (EU) 2020/1828 (Text with EEA relevance). http://data.europa.eu/eli/reg/2023/2854/oj

Data Governance Act. (2022). Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European data governance and amending Regulation (EU) 2018/1724 (Text with EEA relevance). http://data.europa.eu/eli/reg/2022/868/oj

Duchek, S. (2020). Organizational resilience: A capability-based conceptualization. Business Research, 13(1), 215–246. https://doi.org/10.1007/s40685-019-0085-7

Ellström, D., Holtström. J., Berg, E., Josefsson, C. (2022). Dynamic capabilities for digital transformation. *Journal of* Strategy and Management, 15(2), 272–286. doi: https://doi.org/10.1108/JSMA-04-2021-0089

Falkner, G., Heidebrecht, S., Obendiek, A., & Seidl, T. (2024). Digital sovereignty—Rhetoric and reality. Journal of European Public Policy, 31(8), 2099–2120. https://doi.org/10.1080/13501763.2024.2358984

Gill, M., & VanBoskirk, S. (2016). The Digital Maturity Model 4.0 | Forrester. Forrester. https://www.forrester.com/report/the-digital-maturity-model-40/RES131801

Gonzalez-Varona, J. M., Lopez-Paredes, A., Poza, D., & Acebes, F. (2021). Building and development of an organizational competence for digital transformation in SMEs. Journal of Industrial Engineering and Management, 14(1), 15. https://doi.org/10.3926/jiem.3279

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design Science in Information Systems Research. Management Information Systems Quarterly, 28(1), 75–106. https://doi.org/10.2307/25148625

Hillmann, J., & Guenther, E. (2021). Organizational Resilience: A Valuable Construct for Management Research? International Journal of Management Reviews, 23(1), 7–44. https://doi.org/10.1111/ijmr.12239

Kljajić Borštnar, M., & Pucihar, A. (2021). Multi-Attribute Assessment of Digital Maturity of SMEs. Electronics, 10(8). https://doi.org/10.3390/electronics10080885

Ko, A., Mitev, A., Kovacs, T., Feher, P., & Szabo, Z. (2023). Digital Agility, Digital Competitiveness, and Innovative Performance of SMEs. Journal of Competitiveness, 14(4). https://doi.org/10.7441/joc.2022.04.05

Marolt, M., Lenart, G., Kljajić Borštnar, M., & Pucihar, A. (2024). Exploring Digital Transformation Journey Among Micro, Small-, and Medium-Sized Enterprises. Systems, 13(1), 1. https://doi.org/10.3390/systems13010001

Merlo, T. R., Fard, F., & Hawamdeh, S. (2025). Cloud Computing's Impact on the Digital Transformation of the Enterprise: A Mixed-Methods Approach. Sustainability, 17(13), 5755. https://doi.org/10.3390/su17135755

Mithas, S., Tafti, A., & Mitchell, W. (2013). How a Firm's Competitive Environment and Digital Strategic Posture Influence Digital Business Strategy1. Management Information Systems Quarterly, 37(2), 511–536. https://doi.org/10.25300/MISQ/2013/37.2.09

OECD. (2023). OECD SME and Entrepreneurship Outlook 2023. OECD Publishing. https://doi.org/10.1787/342b8564-en

Open DMAT—Digital Maturity Assessment tool | European Digital Innovation Hubs Network. https://european-digital-innovation-hubs.ec.europa.eu/open-dma

Saeedikiya, M., Salunke, S., & Kowalkiewicz, M. (2024). Toward a dynamic capability perspective of digital transformation in SMEs: A study of the mobility sector. Journal of Cleaner Production, 439, 140718. https://doi.org/10.1016/j.jclepro.2024.140718

Sagala, G. H., & Őri, D. (2026). Resilience and digital transformation: A small to medium enterprise resilience framework. Management Decision, 64(13), 126–161. https://doi.org/10.1108/MD-05-2025-1470

Sawang, S., & Sornlertlamvanich, V. (2026). Artificial Intelligence (AI) Maturity in Small and Medium-Sized Enterprises: A Framework of Internalized and Ecosystem-Embedded Capabilities (arXiv:2603.08728). arXiv. https://doi.org/10.48550/arXiv.2603.08728

Schumacher, A., Erol, S., & Sihn, W. (2016). A Maturity Model for Assessing Industry 4.0 Readiness and Maturity of Manufacturing Enterprises. Procedia CIRP, The Sixth International Conference on Changeable, Agile, Reconfigurable and Virtual Production (CARV2016), 52, 161–166. https://doi.org/10.1016/j.procir.2016.07.040

Teece, D. J. (2007). Explicating dynamic capabilities: The nature and microfoundations of (sustainable) enterprise performance. Strategic Management Journal, 28(13), 1319–1350. https://doi.org/10.1002/smj.640

Vial, G. (2019). Understanding digital transformation: A review and a research agenda. The Journal of Strategic Information Systems, 28(2), 118–144. https://doi.org/10.1016/j.jsis.2019.01.003

Warner, K. S. R., & Wäger, M. (2019). Building dynamic capabilities for digital transformation: An ongoing process of strategic renewal. Long Range Planning, 52(3), 326–349. https://doi.org/10.1016/j.lrp.2018.12.001

DIGITAL MATURITY OF SMEs AS A CATALYST FOR SUSTAINABLE DIGITAL FUTURES: A MULTI-CRITERIA ANALYSIS OF EU MEMBER STATES

Eva Ardielli

VSB-Technical University of Ostrava
Faculty of Economics
eva.ardielli@vsb.cz

DOI: 10.35011/IDIMT-2026-99

Keywords

Artificial Intelligence; DESI index, digital transformation; SMEs; twin transition.

Abstract

This paper is based on the analysis of digital transformation of small and medium enterprises (SMEs) in the context of the vision of Sustainable Digital Futures, as a state where the digital infrastructure of small and medium enterprises enables full transparency of the product life cycle and minimization of the energy intensity of processes. SMEs form the backbone of the European economy, but their role in the so-called "Twin Transition" – the synergistic connection of digital and green transformations – is often limited by structural barriers and unequal access to advanced technologies. The research aims to assess and compare the level of digital maturity of SMEs in EU Member States using a multi-criteria weighted sum method (WSA). The analysis is based on current DESI data and the Digital Decade monitoring, focusing on six key indicators, including the adoption of artificial intelligence, big data analytics, and cloud services. The results indicate a significant digital divide between regional leaders and countries with low digital intensity of SMEs, which poses a risk to the inclusiveness of the Union's sustainable development. The findings provide valuable implications for policy makers seeking to overcome the two-speed digital economy in Europe.

1. Digital transformation and the Twin Transition concept

Digital transformation (DT) is understood in the current scientific discourse as a continuum that evolves from the simple digitization of processes, through data-driven decision-making and the development of platform economies, to a fundamental redefinition of business models and governance structures (Azzalin et al., 2024). This transformation is not just a technological change, but a total reorientation of organizations towards network effects and digital data (Veber et al., 2023). In the European context, DT is inextricably linked to the Twin Transition concept, which seeks to synergize digital technologies and green goals (Kovacic et al., 2024). This duality is anchored in key EU policy frameworks, such as the European Green Deal, the Digital Decade, and the Next Generation EU (NGEU), which together aim to increase resilience and climate neutrality through technological modernization (D'Adamo et al., 2025; Verdolini, 2023).

The realization of sustainability goals within the Twin Transition is conditional on the existence of a robust and interoperable data infrastructure (Coors & Padsala, 2024). Advanced technologies such as Digital Twins, Artificial Intelligence (AI), and Big Data require high-quality datasets and transparent data management to fulfil their potential in resource optimization and predictive maintenance (Jurmu et al., 2023). In this sense, Digital Product Passports (DPPs) are a key tool for the circular economy, enabling product lifecycle traceability and efficient recycling (Lang et al., 2024). However, without secure data spaces and legal frameworks for data access, digital innovations risk failing to adequately support environmental ambitions (Verdolini, 2023). The integration of digital tools into the management of cultural heritage or smart cities further demonstrates the need for open and interoperable platforms for sustainable development (Turillazzi et al., 2021; Komninos, 2022).

1.1. Specifics and barriers for small and medium-sized enterprises

Despite the ambitions of EU policies, there is a deep gap between large corporations and the SME sector in the digital economy (Burinskienė & Nalivaikė, 2024; Ćućić et al., 2023). This gap is defined by unequal access to digital infrastructure, financial constraints, and a differing ability to capture value from data (Rahnama et al., 2022). The literature identifies several key barriers that hold SMEs back:

- Financial limits and ROI - SMEs face high initial investments and often uncertain return on investment (ROI), which, together with limited cash flow, increases their risk aversion (Burinskienė & Nalivaikė, 2024; Rollison, 2021),
- Skills and human capital - the lack of digital experts and the low level of digital literacy of management are more critical factors than access to technology itself (Arendt, 2008; Pedone et al., 2023).
- Regulatory burden and standardization - the fragmentation of standards and the complexity of processes, such as BIM implementation in construction, can easily overwhelm small firms with limited resources (Gaffuri et al., 2022; Geoghegan et al., 2023).
- Data sovereignty and cybersecurity concerns about data security and lack of trust in data ecosystems limit the participation of SMEs in broader collaborative networks (Krogt et al., 2023; Kim, 2025).

1.2. The path to an inclusive and sustainable digital future

Targeted support for SMEs is essential to bridge the digital divide and successfully deliver the Twin Transition vision. This should include simplified financing, support for Digital Innovation Hubs (DIHs), and skills development (OECD, 2021). Research highlights that digital innovation and sustainable development are “two sides of the same coin” (Pascale et al., 2024). The key to success is the creation of value constellations in which smaller technology providers collaborate with scientific partners on sustainable solutions (Rahnama et al., 2022). Only through comprehensive support that links technology adoption with environmental metrics (energy efficiency, carbon footprint reduction) can SMEs ensure that they are not left behind in the European economy and actively contribute to building a sustainable future (Niavis et al., 2022).

Although the current literature emphasizes the strategic importance of linking digital transformation and sustainability (the so-called Twin Transition), most studies focus on theoretical frameworks (Veber et al., 2023), or on the barriers of large corporations (Krogt et al., 2023). There is a research gap in understanding how the quantifiable level of digital intensity of SMEs (measured by the DESI index) actually conditions their ability to participate in sustainable digital futures. There is a lack of

deeper empirical analysis that would show, using current data from recent years (2024–2025), whether the gap between "digital leaders" and "laggards" in the segment of small companies is widening in the context of the adoption of technologies critical to sustainability, such as AI and big data analytics.

The main objective of this contribution is to analyze the current state of digitization of SMEs in the EU based on DESI index data and to evaluate to what extent this level of digitization creates the prerequisites for fulfilling the vision of Sustainable Digital Futures (concept that connects digital transformation with environmental and social sustainability). To achieve the stated objective, the following research questions were formulated.

- *RQ1: What is the current state of adoption of AI by SMEs in the EU compared to the strategic objectives of the Digital Decade?*
- *RQ2: What are the differences in the digital maturity of SMEs between selected EU regions and how do these differences affect the inclusiveness of Europe's sustainable digital future?*

2. Methodology

The methodological part is focused on describing the procedure for quantitative analysis of the digital maturity of small and medium-sized enterprises (SMEs) in EU Member States. The evaluation will use a multi-criteria decision-making method, specifically the Weighted Sum Approach (WSA), which allows for an objective comparison of entities based on multiple criteria. For the purposes of DESI index analysis, this method is ideal because it will allow us to aggregate various indicators of SME digital maturity into one comparable score for individual EU countries.

2.1. Data source and selection criteria

The basic data source for the analysis is the Digital Economy and Society Index (DESI) database, based on the current Digital Decade monitoring provided by the European Commission (European Commission, 2026). Specific criteria relevant to the SME sector were selected for research purposes.

- KA1: Artificial intelligence (enterprises using any AI technology) - generative AI (LLM) and predictive models, % of enterprises, 2025.
- KA2: Cloud - cloud computing (enterprises using at least one intermediate or sophisticated cloud computing service such as finance or accounting software applications, enterprise resource planning software applications, customer relationship management software applications or security software applications), % of enterprises, 2024.
- KA3: Data analytics (enterprises performing data analytics) - use of technologies, techniques, or software tools to analyse data to extract patterns, trends, and insights to make conclusions, predictions, and better decision-making with the aim of improving performance, % of enterprises, 2024.
- KA4: Electronic information sharing (enterprises who have used an enterprise resource planning software package to share information between different functional areas, % of enterprises, 2024.
- KA5: SMEs with at least a basic level of digital intensity (basic level requires the usage of at least 4 technologies out of 12 selected technologies), % of enterprises, 2025.
- KA6: SMEs selling online (at least 1% of turnover), % of enterprises, 2025.

All indicators analyzed are related to the category of small businesses (10–249 employees) in all manufacturing and service sectors, excluding the financial sector.

2.2. Weighted sum method

The WSA method is one of the basic methods of multi-criteria evaluation of variants and is based on the principle of linearity and additivity of utility (Brožová et al., 2003); (Ardielli and Bečica, 2018). Given the nature of the input data, where all criteria have a maximization character, the following procedure was chosen:

- Creation of a normalised criterion matrix $R = (r_{ij})$; see formula (1):

$$r_{ij} = \frac{y_{ij} - d_j}{h_j - d_j} \tag{1}$$

  The matrix R represents the matrix of utility values of the i-th variant according to the j-th criterion, where $r_{ij} \in \langle 0; 1 \rangle$.

- Aggregation and utility calculation according to formula (2):

$$u(a_i) = \sum_{j=1}^{n} v_j r_{ij} \tag{2}$$

Based on the resulting values of $u(a_i)$, the EU countries are ranked according to the decreasing value of the utility function to assess the digital maturity reached. This benchmarking will allow to identify structural differences in the digital readiness of SMEs across the EU and interpret them in the context of research questions. For the purposes of this research, the weights are determined using a paired comparison method based on expert estimation (a panel of three academic digital economy experts) with an emphasis on the strategic goals of the Digital Decade 2030 (European Commission, 2026); see Table 1.

Table 1. Weights of the indicators

Indicator	Characteristics	Weight
KA1	Artificial Intelligence	0.25
KA2	Cloud	0.15
KA3	Data Analytics	0.20
KA4	Electronic information sharing	0.10
KA5	Basic Digital Intensity	0.20
KA6	SMEs selling online	0.10

Source: own processing

A higher weight is assigned to the adoption of AI and Big Data, which are identified in the literature as the key to sustainable transformation (Azzalin et al., 2024). The priority of AI (0.25) comes from its role as a general-purpose technology, while the low weight of online sales (0.10) reflects the fact that it is a common commodity capability, not a transformative innovation. To verify robustness, an indicative sensitivity analysis (change in weights by ±10%) was performed, which confirmed the stability of the ranking at the top and bottom positions of the ranking, which reduces the risk of subjective bias.

3. Results and discussion

This part of the paper presents the results of an analysis of the digital maturity of small and medium-sized enterprises (SMEs) in EU Member States, obtained by applying the WSA to selected DESI indicators. The resulting $u(a_i)$ values reflect not only the overall level of digitalization, but also the readiness of the SME sector to integrate advanced technologies (AI, Big Data), which are key to fulfilling the vision of a sustainable digital future. The dominance of digital leaders in the ranking is partly determined by the weighting that favours transformative technologies (AI, Big Data) over ordinary digitalization. However, this approach is deliberate and reflects the strategic priorities of the EU. Another limitation is that DESI indicators serve as proxy variables for digital maturity but do not reflect the real degree of integration of these technologies into sustainable business strategies.

Table 2 presents the final ranking of the digital maturity of SME in EU Member States. The top countries not only demonstrate high Basic Digital Intensity (KA5) but also dominate in AI integration. This fact suggests that these economies have already moved beyond the stage of simply digitizing processes and are moving towards predictive management models that are essential for an effective Twin Transition (Lang et al., 2024). The results confirm the deep Digital Divide between North-West and South-East Europe. SMEs in countries at the bottom of the ranking show critically low values for criterion KA1 (AI), which, according to research findings, means that SMEs without the ability to integrate into digital ecosystems (requiring interoperability for AI analytics) risk being excluded from sustainable supply chains.

Table 2. Evaluation of EU countries using the WSA method (2025)

Final order	Variant	Total utility value	Final order	Variant	Total utility value
1	Denmark	0.94866	15	Portugal	0.42578
2	Finland	0.82969	16	Cyprus	0.41592
3	Sweden	0.78238	17	Hungary	0.41066
4	The Netherlands	0.76872	18	Estonia	0.40847
5	Belgium	0.76001	19	Slovenia	0.38891
6	Malta	0.63400	20	Italy	0.36703
7	Germany	0.56279	21	France	0.33920
8	Ireland	0.53956	22	Latvia	0.29872
9	Croatia	0.49636	23	Czechia	0.29198
10	Luxembourg	0.48192	24	Slovakia	0.26741
11	Spain	0.47608	25	Poland	0.24285
12	Austria	0.45967	26	Greece	0.23708
13	European Union	0.44543	27	Romania	0.13311
14	Lithuania	0.43939	28	Bulgaria	0.07113

Source: based on data from European Commission (2026), own calculations

A key indicator of the general state of SME digitalization in the European Union is the calculated average value of the $u(a_i)$ score (EU average), which is 0.44543. This value serves as a dividing line between digitally progressive economies and countries that face structural difficulties in adopting advanced technologies. Countries exceeding the value of 0.44543 (especially the Scandinavian

countries, Benelux, and selected Western economies) show high homogeneity in fulfilling all six monitored criteria. The identified regional disparities in the digital maturity of SMEs correlate with previous analyses of digital processes in the EU27. As reported by Doucek et al. (2025), a group of leaders led by Denmark, Finland, and Sweden show a consistent lead in the adoption of both AI and ERP systems, while countries such as Bulgaria, Romania, and the Czech Republic remain at the opposite pole in the process of digitalizing processes and implementing new technologies into practice (Doucek et al., 2025).

Countries with scores below 0.44543 show considerable heterogeneity. The biggest drop compared to the EU average is seen in criteria KA1 (AI) and KA3 (Data analytics). Even in cases where these countries achieve solid results in e-commerce (KA6) or basic digital intensity (KA5), the absence of advanced analytics tools brings their overall score below the EU average. These findings confirm the risk of "digital lagging behind", where below-average countries may struggle with future interoperability within pan-European data spaces (Jurmu et al., 2023).

The fact that the average value does not even reach half of the theoretical maximum (Ui=1.0) indicates a significant untapped potential of the SME sector. For the sustainable transformation of the EU, the wide dispersion between the best and worst results is particularly alarming, indicating that EU support for digitalization needs to be much more differentiated and targeted at specific barriers in below-average regions (Rollison, 2021).

The results of the analysis indicate that the identified regional Digital Divide is not only a technological phenomenon, but is transforming into a deeper structural problem, which can be defined as the Sustainability Gap. Although digital leaders use the advanced maturity of SMEs to accelerate the Twin Transition through predictive models and data integration, below-average countries (score Ui< 0.44543) show the risk of technological lock-in in less sophisticated processes. From a scientific point of view, these findings go beyond the descriptive nature of the DESI index by identifying a causality between the absence of analytical capabilities of SMEs and their future inability to participate in the inclusive development of sustainable digital ecosystems in the EU. Thus, the digital maturity of SMEs ceases to be a mere measure of competitiveness and becomes a fundamental prerequisite for the ecological resilience and social inclusiveness of the European economic area.

4. Conclusion

The paper analyzed the digital maturity of SMEs in the EU as a key prerequisite to realize a sustainable digital future. Through multi-criteria WSA analysis applied to DESI index data, it was possible to quantify the current state of SME digitalization and identify critical points that prevent the effective connection of technological development with environmental goals.

The research yields key conclusions that directly answer the stated research questions. In connection with RQ1 the results show that despite the EU's declared ambitions, the real adoption of AI in the SME sector is still at an early stage. The EU average score (0.44543) confirms that most businesses use digital tools for operational activities (cloud, e-commerce) rather than for data-driven strategic transformation (AI, big data). This gap represents a major barrier to the implementation of comprehensive sustainability solutions such as digital product passports or predictive resource optimisation. Achieving the EU Digital Decade target of 75% AI adoption appears highly unlikely; in 2024, only 13.48% of all enterprises utilized at least one AI technology, while the average for SMEs lagged even further behind at 12.64%. In connection with RQ2 the analysis confirmed the existence of a digital divide between Member States, especially countries with high digital intensity (Denmark, Finland, Sweden, the Netherlands and Belgium and countries with above-average results (score Ui> 0.44543) and countries with a large lag in digitalization (Slovakia, Poland, Greece,

Romania and Bulgaria). The Czech Republic is ranked 22nd among EU countries. The gap between digital leaders and countries below the EU average indicates the emergence of a two-speed sustainable Europe. Regarding the inclusiveness of a sustainable future, there is a potential concern that digitally lagging SMEs could face challenges in maintaining their positions within global supply chains. This is mainly due to potential difficulties in meeting emerging requirements for data interoperability and sustainability reporting.

In conclusion, the digital transformation of small and medium enterprises is to seen as an essential basis for the Twin Transition. Achieving sustainable digital futures depends on reducing regional disparities and facilitating the shift from basic digitalization toward advanced data analytics and AI. Policy efforts should prioritize Digital Innovation Hubs (DIHs) to mitigate barriers such as high investment costs and the know-how gap in underperforming regions.

Acknowledgement

This paper was created with the financial support of project No. CZ.11.3.119/0.0/0.0/16_013/0003093 "Increasing the knowledge and skills of university students in the field of implementation of public policies in the Czech Republic and Poland and improving their applicability in the labor market" (Program Interreg V-A Czech Republic-Poland) and with the financial support of project number 22430108 “Inventory of Good Practices in Local Housing Policy of V4 Countries Supported by the Visegrad Fund.” The project is co-financed by the governments of Czechia, Hungary, Poland and Slovakia through Visegrad Grants from the International Visegrad Fund. The mission of the fund is to advance ideas for sustainable regional cooperation in Central Europe.

References

Ardielli, E., & Bečica, J. (2018). Multi-criteria Evaluation of the State of Professional Theatres in the Czech Republic in Terms of Mixed Public Goods Provided to the Citizens. Review of Economic Perspectives, 18(2), 155-176. https://doi.org/10.2478/revecp-2018-0009

Arendt, Ł. (2008). Barriers to ICT adoption in SMEs: how to bridge the digital divide? Journal of Systems and Information Technology, 10(2), 93–108. https://doi.org/10.1108/13287260810897738

Azzalin, M., Lauria, M., Gulletta, A., & Melchini, T. (2024). New paradigms shift in buildings: experimental application of Digital Twin for safety and well-being. IOP Conference Series: Earth and Environmental Science, 1402(1), 012046. https://doi.org/10.1088/1755-1315/1402/1/012046

Brožová, H., Houška, M., & Šubrt, T. (2003). Modely pro vícekriteriální rozhodování, Praha, Credit.

Burinskienė, A., & Nalivaikė, J. (2024). Digital and Sustainable (Twin) Transformations: A Case of SMEs in the European Union. Sustainability, 16(4), 1533. https://doi.org/10.3390/su16041533

Coors, V., & Padsala, R. (2024). Urban Digital Twins empowering energy transition: Citizen-driven sustainable urban transformation towards positive energy districts. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLVIII-4/W10-2024, 51–56. https://doi.org/10.5194/isprs-archives-xlviii-4-w10-2024-51-2024

Ćućić, D., Pancić, M., & Serdarušić, H. (2023). The financial and social impacts of the approved projects in the area of digital and green transition on the Croatian SMEs. EU and Comparative Law Issues and Challenges Series (ECLIC), 7, 223–242. https://doi.org/10.25234/eclic/27464

D’Adamo, I., Fratocchi, L., Grosso, C., & Tavana, M. (2025). An integrated business strategy for the twin transition: Leveraging digital product passports and circular economy models. Business Strategy and the Environment, 34(7), 9008–9022. https://doi.org/10.1002/bse.70065

Doucek, P., Nedomova, L., & Maryska, M. (2025). Digitalization in Business. Proceedings of the 33rd IDIMT Conference, 205-214. https://doi.org/10.35011/IDIMT-2025-205

European Commission. (2026). Digital Economy and Society Index (DESI). https://digital-strategy.ec.europa.eu/en/policies/desi

Gaffuri, G., Chartier, I., Ribeiro, P. D., Nastran, S., Volpe, M., Dor, I., Murillo, G., Chanson, E., Gonzalez, I., & Hofer, M. (2022). Supporting innovation in energy sector: The SMEs role and DIGIFED success cases. Proceedings of the 17th European Conference on Innovation and Entrepreneurship (ECIE), 232–239. https://doi.org/10.34190/ecie.17.1.458

Geoghegan, H. J., Jensen, F. W., Kershaw, T., & Codinhoto, R. (2023). Innovation realisation for digitalising Dutch small architecture practices: state of the art review. Proceedings of the Institution of Civil Engineers - Management, Procurement and Law, 176(4), 176-191. https://doi.org/10.1680/jmapl.22.00018

Jurmu, M., Niskanen, I., Kinnula, A., Kääriäinen, J., Ylikerälä, M., Räsänen, P., & Tuikka, T. (2023). Exploring the role of federated data spaces in implementing twin transition within manufacturing ecosystems. Sensors, 23(9), 4315. https://doi.org/10.3390/s23094315

Kim, J. (2025). Digital twins as planetary media in Europe and China: Optimization and resilience in scientific and socioeconomic domains. New Media & Society, 27(8), 4515-4532. https://doi.org/10.1177/14614448251336434

Komninos, N. (2022). Transformation of Industry Ecosystems in Cities and Regions: A Generic Pathway for Smart and Green Transition. Sustainability, 14(15), 9694. https://doi.org/10.3390/su14159694

Kovacic, Z., Casañas, C. G., Argüelles, L., Serrano, P. Y., Ribera-Fumaz, R., Prause, L., & March, H. (2024). The twin green and digital transition: High-level policy or science fiction? Environment and Planning E: Nature and Space, 7(6), 2251–2278. https://doi.org/10.1177/25148486241258046

Krogt, A. V. der, Rensma, A., & Reijswoud, V. van. (2023). Twin green and digital innovation by SMEs in the construction sector. Proceedings of the 18th European Conference on Innovation and Entrepreneurship (ECIE), 18(2), 891-898. https://doi.org/10.34190/ecie.18.2.1473

Lang, S., Müller, M. M., & Stöber, L. F. (2024). Twin Transition – Hidden Links between the Green and Digital Transition. Journal of Innovation Economics & Management, 45(3), 57-94. https://doi.org/10.3917/jie.045.0057

Niavis, H., Laskari, M., & Fergadiotou, I. (2022). Trusted DBL: A blockchain-based digital twin for sustainable and interoperable building performance evaluation. 2022 7th International Conference on Smart and Sustainable Technologies (SpliTech), 1-6. https://doi.org/10.23919/splitech55088.2022.9854287

OECD. (2021). No net zero without SMEs. OECD SME and Entrepreneurship Papers. https://doi.org/10.1787/bab63915-en

Pascale, G. D., Faccilongo, N., Riefolo, M., Romagno, A., & Silvestri, R. (2024). Digital innovation and sustainable development: Two sides of the same coin. In Sustainable Development - Glocal Perspective. IntechOpen. https://doi.org/10.5772/intechopen.112294

Pedone, A., Angotti, R., & Barricelli, D. (2023). Training and skills development in SMEs. Sinappsi, 13(1), 32–49. https://doi.org/10.53223/sinappsi_2023-01-3

Rahnama, H., Johansen, K., Larsson, L., & Rönnbäck, A. Ö. (2022). Collaboration in value constellations for sustainable production: The perspective of small technology solution providers. Sustainability, 14(8), 4794. https://doi.org/10.3390/su14084794

Rollison, J. (2021). SME digitalisation to “Build Back Better”. OECD SME and Entrepreneurship Papers. https://doi.org/10.1787/50193089-en

Turillazzi, B., Leoni, G., Gaspari, J., Massari, M., & Boulanger, S. O. M. (2021). Cultural heritage and digital tools: The rock interoperable platform. International Journal of Environmental Impacts, 4(3), 276-288. https://doi.org/10.2495/ei-v4-n3-276-288

Veber, J. et al. (2023). Udržitelnost a udržitelný management. Praha, Grada.

Verdolini, V. (2023). Interlinkages between the just ecological transition and the digital transformation, ETUI, Brussels.

MARKET DYNAMICS AND RISKS IN THE GLOBAL MACHINE LEARNING INDUSTRY

Maria Bartekova

Bratislava University of Economics and Business, Slovakia
Faculty of Business Management
maria.bartekova@euba.sk

DOI: 10.35011/IDIMT-2026-107

Keywords

Artificial Intelligence; machine learning; market growth; systemic risk.

Abstract

The global machine learning (ML) industry has emerged as a key driver of digital transformation and economic development, expanding rapidly across multiple sectors. This paper examines the relationship between the growth of the machine learning market and the economic performance of European countries, with a focus on the real GDP growth rate. Using data from Statista and Eurostat for the period 2020–2031, the study applies correlation and regression analysis to identify potential dependencies between ML market expansion and macroeconomic indicators. The findings suggest that while the ML market demonstrates strong and consistent growth, its relationship with short-term economic indicators such as GDP growth is not always statistically significant. This indicates that the expansion of machine learning is influenced more by structural factors, including digital infrastructure, technological readiness, and investment capacity, rather than short-term economic fluctuations. The results highlight the importance of long-term economic and technological conditions for sustaining growth in the ML industry.

1. Introduction

The global machine learning (ML) industry has become a central driver of technological transformation and economic growth, expanding across sectors such as healthcare, finance, retail, transportation, and manufacturing. Industry reports project substantial market growth over the coming decade, reflecting the increasing integration of ML technologies into enterprise systems and consumer applications (The Business Research Company, 2024; Grand View Research, 2023).

This growth is driven by advances in computational power, large-scale data availability, cloud infrastructure, and algorithmic innovation (Grand View Research, 2023). Demand for automation, predictive analytics, and intelligent decision-support systems has intensified competition among technology firms, especially cloud service providers and multinational technology corporations. At the same time, high market concentration creates barriers to entry while accelerating innovation cycles (Technavio, 2023).

Despite rapid development, the ML industry operates within a complex risk environment. Ethical concerns related to algorithmic bias, transparency, and accountability continue to attract regulatory

scrutiny (Dignum, 2021). ML systems also create operational and systemic risks, including model failures, data vulnerabilities, and unintended impacts in critical sectors (Olutimehin et al., 2024). Further risks stem from infrastructure constraints, including dependence on specialized semiconductor supply chains and large-scale data centers.

Although existing literature discusses market growth, technical progress, and ethical risks, fewer studies integrate these perspectives into a unified framework analysing both market dynamics and systemic risk exposure. Most research focuses separately on market size and adoption trends, technical performance, or ethical governance (The Business Research Company, 2024; Grand View Research, 2023; Dignum, 2021).

This study addresses this gap by examining the interrelationship between market dynamics and systemic risk factors in the global ML industry, providing a broader perspective for policymakers, investors, and industry stakeholders.

2. Literature review

2.1. Growth and market expansion of Machine Learning

The machine learning (ML) market has experienced rapid growth and has become a key component of the digital economy, driven by increasing data availability, advances in computational power, and the expansion of cloud infrastructure (Jordan & Mitchell, 2015; Rahman, 2024). ML technologies are widely applied across sectors such as finance, healthcare, retail, and logistics, contributing to productivity and innovation.

The literature highlights that market expansion is primarily driven by demand for automation, predictive analytics, and data-driven decision-making. Companies increasingly integrate ML into core processes, enhancing efficiency and competitiveness (Yesuf & Fields, 2025). Digital transformation further accelerates this trend by enabling firms to better utilize data, personalize services, and respond to market changes (Olutimehin et al., 2024; Hasan, 2025; Naaman, Ahmed & Zeebaree, 2025).

However, growth is uneven and concentrated in technologically advanced economies and large firms with access to data, capital, and infrastructure. This leads to higher market concentration and barriers to entry (Crémer, de Montjoye & Schweitzer, 2019). At the same time, challenges such as high implementation costs, lack of skilled labour, and regulatory uncertainty may limit further expansion (Benga & Elhamma, 2024; Dignum, 2021).

Overall, it can be stated that the literature perceives the growth and market expansion of machine learning as the result of a combination of technological progress, growing business demand and a deeper digital transformation of the economy. However, it also points out that this growth is unevenly distributed, concentrated in the hands of large players and conditioned by the availability of data, infrastructure, talent and an appropriate regulatory framework. That is why, when examining the machine learning market, it is necessary to analyse not only its growth potential, but also structural risks that may affect its long-term stability.

2.2. Technological and operational risks

The literature on machine learning risks focuses mainly on technical reliability, model robustness, and unintended consequences of deployment. As ML systems are increasingly used in critical sectors such as finance, healthcare, transport, and cybersecurity, technical failures may have broader systemic effects rather than only isolated consequences (Dignum, 2021; Olutimehin et al., 2024).

One key risk concerns model reliability. ML models may perform poorly when real-world data differ from training data, a problem known as dataset shift or concept drift. This can lead to incorrect predictions, especially in high-stakes areas such as medical diagnostics, credit scoring, or risk assessment. Another important risk is cybersecurity. ML systems can be exposed to adversarial attacks, data poisoning, or manipulation of inputs, which may undermine their accuracy and trustworthiness (Jordan & Mitchell, 2015; Dignum, 2021).

Beyond technical risks, algorithmic bias represents a major ethical and operational challenge. Bias may arise from unrepresentative data, historical inequalities, or inappropriate model design. In areas such as recruitment, loan approval, healthcare, or policing, biased outputs may reinforce discrimination and expose organizations to legal, reputational, and regulatory risks (Dignum, 2021; Olutimehin et al., 2024).

Operational scalability further increases risk. When ML systems are deployed globally, even small errors can affect millions of users or multiple sectors at the same time. This is particularly relevant when many organizations rely on similar models, cloud providers, datasets, or software infrastructures. Such concentration may create a "cluster problem", where failure in one critical component can spread across interconnected systems (Crémer, de Montjoye & Schweitzer, 2019; Technavio, 2023).

The relationship between market dynamics and risk is therefore two-sided. Rapid growth, investment, and adoption create economic opportunities, but they may also increase systemic vulnerability when ML development is concentrated among a small number of firms, regions, or infrastructures. For this reason, ML market analysis should consider not only growth potential, but also technological dependency, concentration, regulatory uncertainty, and operational fragility (Grand View Research, 2023; The Business Research Company, 2024; Benga & Elhamma, 2024).

3. Data and methodology

The aim of this study is to examine the relationship between the growth of the machine learning (ML) market and the economic performance of European countries, with a particular focus on the Real GDP growth rate.

3.1. Data source

The empirical analysis is based on secondary data obtained from the Statista database and complementary macroeconomic indicators from Eurostat. The dataset includes information on the development of the machine learning market in Europe as well as macroeconomic indicators for selected European countries.

The following variables were used:

- Machine Learning Market Size (ML_MS) – measured as the estimated annual revenue generated by the machine learning market, expressed in billions of USD. In this paper, market size does not refer to the number of ML services sold, the stock market value of ML firms, or the amount of venture capital invested. It represents the monetary value of the ML market as reported in Statista market estimates.
- Machine Learning Market Growth (ML_GR) – measured as the annual percentage change in ML market revenue. It captures the rate at which the market expands over time. Therefore, growth is interpreted as revenue growth, not as the number of adopting firms or the number of deployed ML systems.

- Real GDP Growth Rate (GDP_GR) – measured as the annual percentage change in real GDP, obtained from Eurostat, representing short-term macroeconomic performance across European countries. This variable is used only as a macroeconomic comparator and not as a direct proxy for firm-level ML adoption or production.

The dataset combines time-series and cross-sectional elements, allowing for both temporal and comparative analysis across countries. Prior to analysis, the data were cleaned, aligned across time periods, and, where necessary, normalized to ensure comparability between countries with different economic sizes.

3.1.1. Derivation of market growth and adoption indicators

The claims regarding growth, adoption, and market dynamics are not inferred from monetary market size alone. They are derived from three complementary indicators: (1) annual market growth rates of the machine learning industry, (2) total machine learning market size expressed in monetary terms, and (3) the regional distribution of machine learning market revenues. Market growth is therefore interpreted as a combination of percentage expansion and increasing market value, while market dynamics are understood as changes in the scale, structure, geographical concentration, and technological composition of the industry.

Machine Learning Adoption is not directly measured at the firm level, because the available dataset does not include individual company-level information on whether and how firms implemented ML solutions. The paper also does not distinguish between companies that produce ML technologies and companies that use them. This is an important limitation, because a country may generate market revenue from exporting ML technologies without necessarily having broad domestic adoption, while another country may use ML extensively but not produce market revenue at the same scale. Therefore, adoption is interpreted cautiously through market-level diffusion, reflected in rising market revenue, positive annual growth rates, and the increasing commercial relevance of ML technological components.

Other explanatory factors, such as digital infrastructure, technological readiness, investment capacity, cloud computing infrastructure, semiconductor availability, skilled labour, and regulatory conditions, are identified from the literature as structural conditions that may shape ML market expansion. These factors are not directly included as regression variables because comparable data were not available for all observations in the dataset. Consequently, they are used as interpretive explanatory factors rather than as statistically tested predictors. This distinction is important because the empirical results suggest that short-term GDP growth alone does not sufficiently explain ML market development.

3.1.2. Sample selection and limitations

The European regression model is based on countries for which comparable data were simultaneously available for ML market indicators and real GDP growth. The analysis does not use firm-level adoption data and does not distinguish between companies producing ML technologies and companies using ML technologies. The ML market data capture market revenue attributed to a given country or region, while the GDP growth variable captures macroeconomic performance. Therefore, the model does not directly measure whether firms in a country adopt ML solutions, nor does it measure whether domestic firms produce ML technologies for export.

For this reason, the regression analysis should be interpreted cautiously. The relationship tested in the model is not a direct causal link between firm-level ML adoption and GDP growth. Rather, it is an exploratory comparison of macroeconomic performance and market-level ML indicators in selected European countries. The analysis is intended to indicate whether countries with stronger short-term

GDP growth also show stronger ML market indicators, but it cannot identify the exact mechanisms of production, adoption, export orientation, or company-level implementation.

Similarly, the worldwide component model is based on a limited set of comparable segment-level observations for the main technological components of the global machine learning market. The sample size of N = 6 reflects the restricted number of consistent observations available in the secondary dataset. This is sufficient only for an exploratory description of market composition and should not be interpreted as robust causal evidence. The model is therefore used to discuss the relative development of technological components rather than to prove that these components independently determine total market size.

3.2. Hypothesis formulation

Based on the preliminary analysis and theoretical framework, the following hypothesis was established:

- H_1: There is an observable association between real GDP growth and machine learning market indicators in selected European countries.

To analyse the relationship between the machine learning market and economic growth, both descriptive and inferential statistical methods were applied.

In the first step, descriptive statistics were used to examine the basic characteristics of the dataset, including mean values, standard deviation, and distribution properties. Special attention was given to skewness and kurtosis indicators to assess the normality of the data distribution. Subsequently, correlation analysis was conducted to identify the strength and direction of the relationship between ML market indicators and the Real GDP growth rate. Pearson's correlation coefficient was used in cases where the assumption of normality was satisfied; otherwise, Spearman's rank correlation coefficient was applied. In addition, countries were grouped based on economic performance levels (e.g., higher-growth vs. lower-growth economies), and non-parametric tests such as the Kruskal–Wallis test were applied to examine differences between groups where normality assumptions were not met.

4. Results and discussion

Figure 1 illustrates the global development of the machine learning industry over time. The trend demonstrates a steady and dynamic increase in revenue growth, indicating rapid market expansion driven by technological innovation, increased data availability, and growing demand for intelligent systems. This upward trajectory suggests that ML is becoming a more important commercial component of digital transformation across industries. However, the figure should be interpreted as a market-level indicator of revenue growth, not as direct evidence of the number of firms adopting ML technologies.

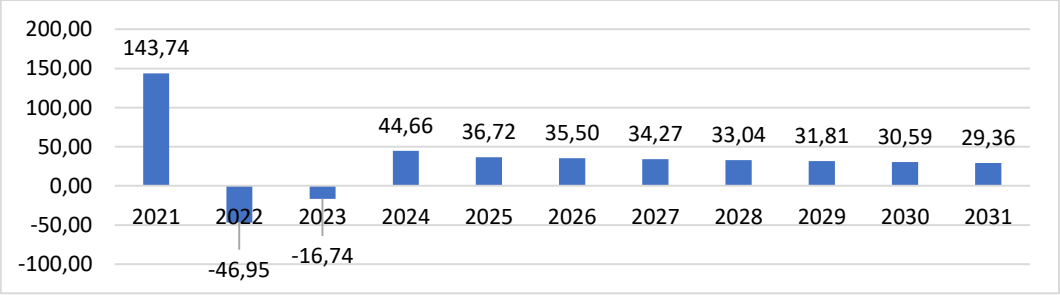


Figure 1. Development of Machine Learning Industry – worldwide (in %) Source: Statista

Figure 2 shows the regional structure of the machine learning industry. The distribution indicates a strong concentration of market activity in technologically advanced regions, particularly North America and parts of Asia. This concentration reflects disparities in access to capital, data infrastructure, skilled labour, technological expertise, and large platform ecosystems. The dominance of certain regions suggests the presence of network effects and economies of scale, which reinforce the competitive advantage of leading markets and may contribute to the cluster problem discussed in the theoretical part.

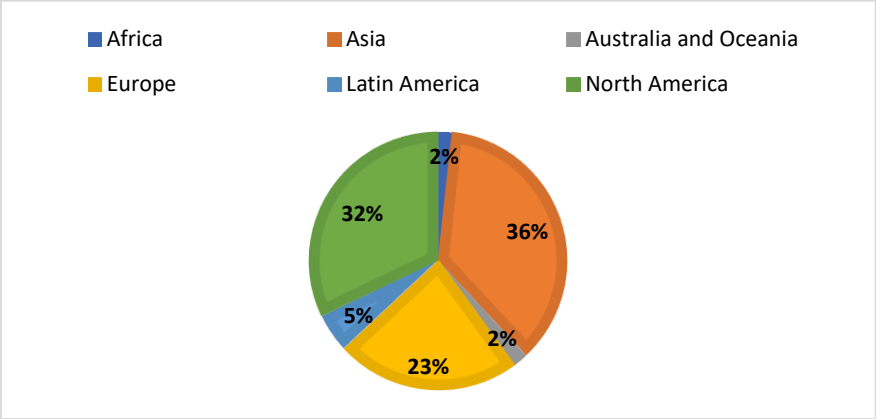


Figure 2. Machine Learning Industry – Regional Structure (in bil. USD, 2025) Source: Statista

Table 1 presents an exploratory linear regression model examining the relationship between real GDP growth and selected ML market indicators in Europe. The model includes ML market growth and ML market size as explanatory variables. However, due to the very small sample size, the results cannot be interpreted as robust statistical evidence. The model is used only to explore whether short-term GDP growth is associated with market-level ML indicators in the available data. The results show that the model is not statistically significant, which suggests that short-term GDP growth does not provide a sufficient explanation of ML market development in the analysed sample. This supports the interpretation that ML expansion is more likely shaped by longer-term structural conditions, such as digital infrastructure, investment capacity, technological readiness, availability of skilled labour, and regulatory environment.

Table 1. Linear regression (Overall Model Test) – Europe

Model	R	R^2	F	df1	df2	p
1	0.716	0.513	1.05	2	2	0.487

Note. Models estimated using sample size of N=5 Source: own calculations according to Statista (2026) and Eurostat (2026)

Table 2 presents the estimated coefficients of the linear regression model examining the relationship between real GDP growth and machine learning market indicators in Europe. The results show that neither ML market growth nor ML market size has a statistically significant effect on real GDP growth, as both p-values are considerably above the conventional 0.05 significance level.

Table 2. Linear regression (Model Coefficients)

Predictor	Estimate	SE	t	p
Intercept	-2.70	6.9519	-0.3877	0.736
ML_GR_E	-0.0009	0.0364	-0.0262	0.981
ML_MS_E	0.0003	-0.0004	0.6760	0.569

Source: own calculations according to Statista (2026) and Eurostat (2026)

5. Conclusion

This study examined the relationship between the growth of the machine learning market and the economic performance of European countries, with a focus on the Real GDP growth rate. The results indicate that the machine learning industry is experiencing strong and continuous market expansion, driven primarily by technological innovation, increasing data availability, and growing demand for intelligent solutions. However, these findings should be interpreted as market-level evidence based on revenue and growth indicators, not as direct firm-level evidence of adoption.

The empirical results do not confirm a statistically significant relationship between real GDP growth and ML market indicators in the analysed European sample. This suggests that short-term macroeconomic growth alone is not sufficient to explain the development of the ML market. The expansion of machine learning appears to be more strongly connected with long-term structural conditions, such as digital infrastructure, technological readiness, innovation capacity, availability of skilled labour, access to capital, and regulatory frameworks.

The analysis of technological components should be interpreted as a descriptive assessment of market structure rather than as causal evidence. Since the global ML market consists of several technological segments, the relationship between the total market and its components is partly mechanical. Nevertheless, the segment-based perspective is useful because it shows which technological domains contribute to the composition and development of the ML industry. Computer Vision, AI Robotics, Autonomous and Sensor Technologies, and Natural Language Processing represent important areas of commercial application, although their development may differ depending on sectoral demand and investment conditions.

Overall, the results indicate that the development of the machine learning industry depends more on long-term structural factors and technological ecosystems than on short-term economic fluctuations. These findings have important implications for policymakers and businesses, emphasizing the need to support technological development, invest in digital infrastructure, reduce dependency risks, strengthen governance mechanisms, and create favourable conditions for innovation in order to sustain growth in the machine learning sector.

Acknowledgement

This research paper is a partial result of the research project VEGA no. 1/0660/26 “Enterprise Risk Management in the Transition to a Circular Economy: Impact on Business Performance and Strategic Sustainability”.

References

Benga, B., & Elhamma, A. (2024). Navigating the digital frontier: A literature review on business digitalization. European Scientific Journal. https://eujournal.org/index.php/esj/article/view/17937

Crémer, J., de Montjoye, Y. A., & Schweitzer, H. (2019). Competition policy for the digital era. European Commission. https://ec.europa.eu/competition/publications/reports/kd0419345enn.pdf

Denmamode, L., & Panchoo, S. (2024). Digital transformation: A roadmap to leverage businesses for SMEs. Journal of Business and Enterprise Development. https://doi.org/10.47963/jobed.v12i.1507

Dignum, V. (2021). The role and challenges of education for responsible AI. Social Sciences, 10(7), 1–14. https://doi.org/10.14324/lre.19.1.01

Eurostat. (2026). Real GDP growth rate. https://ec.europa.eu/eurostat/databrowser/view/tec00115/default/table?lang=en

Grand View Research. (2023). Machine learning market size, share & trends analysis report. Grand View Research. https://www.grandviewresearch.com/industry-analysis/machine-learning-market

Hasan, R. (2025). Enhancing market competitiveness through AI-powered SEO and digital marketing strategies in e-commerce. ASRC Procedia. https://doi.org/10.63125/31tpjc54

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. Science, 349(6245), 255–260. https://doi.org/10.1126/science.aaa8415

Makhfirah, M. N. (2024). Technology-based SMEs development: Big opportunities or new risks? Proceeding of International Students Conference of Economics and Business Excellence. https://doi.org/10.33830/iscebe.v1i1.3638

Naaman, D., Ahmed, B. T., & Zeebaree, S. R. M. (2025). E-business and digital marketing strategies: Innovations, challenges, and emerging trends. Asian Journal of Research in Computer Science. https://doi.org/10.9734/ajrcos/2025/v18i4612

Olutimehin, D. O., Ofodile, O. C., Ejibe, I., Odunaiya, O. G., & Soyombo, O. T. (2024). Implementing AI in business models: Strategies for efficiency and innovation. International Journal of Management & Entrepreneurship Research. https://doi.org/10.51594/ijmer.v6i3.940

Rahman, A. (2024). AI and machine learning in business process automation: Innovating ways AI can enhance operational efficiencies or customer experiences in U. S. enterprises. Non Human Journal. https://doi.org/10.70008/jmldeds.v1i01.41

Statista. (2026). Streaming worldwide - statistics & facts. https://www.statista.com/outlook/tmo/artificial-intelligence/machine-learning/worldwide?srsltid=AfmBOoqtkoJ-8TuSCibfzSTLGGNkBPtWGJa-Ri07sRYFtZ6Ng7sD3Y7i

Technavio. (2023). Machine learning market analysis report 2023–2027. Technavio. https://www.technavio.com/report/machine-learning-market-size-industry-analysis

The Business Research Company. (2024). Machine learning global market report 2024. The Business Research Company. https://www.thebusinessresearchcompany.com/report/machine-learning-global-market-report

The jamovi project (2022). jamovi. (Version 2.3) [Computer Software]. https://www.jamovi.org

Yesuf, Y., & Fields, Z. (2025). Artificial intelligence adoption as a driver of innovation and competitiveness in SMEs: A bibliometric and systematic review. F1000Research. https://doi.org/10.12688/f1000research.171494.1

STRATEGIC ASCENDANCE OF INFORMATION TECHNOLOGY IN BUSINESS STRATEGY AND PERFORMANCE

Martin Zbořil

Prague University of Economics and Business
Faculty of Informatics and Statistics
martin.zboril@vse.cz

DOI: 10.35011/IDIMT-2026-115

Keywords

Information technology evolution; business-IT strategic alignment; digital transformation; IT capabilities; artificial intelligence in business; systematic literature review.

Abstract

Information technology (IT) has undergone a remarkable transformation from a back-office support tool to a central pillar of business strategy and organizational performance. This review article traces the historical trajectory of IT's role in business – from early data processing systems of the 1960s through the rise of enterprise systems, the internet revolution, and the current era of artificial intelligence and cloud computing. Drawing on a systematic literature review of numerous scholarly sources, the paper synthesizes key theoretical perspectives including the productivity paradox, the resource-based view of IT capabilities, and strategic alignment models. An integrative framework is proposed that maps the evolutionary phases of IT's strategic ascendance across three dimensions: economic, organizational, technological. The findings confirm that IT has shifted from an enabler of process automation to a fundamental driver of value creation, innovation, and long-term competitive positioning. The paper concludes with implications for both researchers and practitioners navigating the digitalization landscape.

1. Introduction

The relationship between information technology (IT) and business performance has been widely debated in management and information systems research for over five decades. From early mainframe computers automating payroll to today's AI systems predicting consumer behavior in real time, IT has fundamentally reshaped how organizations create, deliver, and capture value. As Brynjolfsson and Hitt (1998) argue, the question is no longer whether IT contributes to business performance, but how and under what conditions this contribution is maximized. The information revolution, beginning in the early 1970s, introduced Information and Communication Technology (ICT) that brought fundamental changes to how people work, communicate, learn, and interact, profoundly transforming business and government practices (Hannan, 2010). As Toffler (1980) described in The Third Wave, the ICT era introduced in the knowledge economy, enabling vast amounts of data to be stored, processed, and transmitted globally across virtually every economic sector.

The strategic importance of IT has evolved considerably. Porter and Millar (1985) were among the first to articulate how information technology gives companies competitive advantage by transforming the value chain. In the decades that followed, scholars such as Drucker (1963), who predicted the rise of knowledge workers, and Carr (2003), who argued that IT no longer matters as a differentiator, have shaped a rich and sometimes contradictory discourse. The productivity paradox, famously captured by Solow's (1987) observation that computers are everywhere except in the productivity statistics, stimulated an entire generation of research seeking to reconcile massive IT investments with measurable economic returns (Brynjolfsson & Hitt, 2003). Today, the discourse has shifted toward digital transformation, where IT is not merely a tool but a foundational element of business model innovation, ecosystem orchestration, and organizational agility. The convergence of cloud computing, big data analytics, artificial intelligence, and the internet of things has created what Schwab (2016) termed the Fourth Industrial Revolution – a period characterized by the merging of information, communication, and technology with the industrial world. In this context, understanding the historical trajectory and strategic evolution of IT in business is not merely an academic exercise; it is essential for practitioners seeking to navigate an increasingly complex digital landscape.

This paper presents a systematic literature review that traces the strategic ascendance of IT in business strategy and performance. The review is organized chronologically and thematically, examining how IT's role has evolved from operational support to strategic imperative. The contribution of this paper lies in developing an integrative framework that synthesizes over six decades of research into a coherent periodization of IT's strategic evolution across economic, organizational and technological dimensions.

2. Methodology

This study uses a systematic literature review, a well-established method for synthesizing existing knowledge on a topic (Kraus et al., 2020). The review follows recognized guidelines for structured literature reviews in information systems and business research, with the goal of identifying and analyzing scholarly work on the relationship between IT, business strategy, and organizational performance across different historical periods.

The literature search drew on several academic databases, including Web of Science, Scopus, ProQuest, and Google Scholar. Keywords such as "information technology AND business performance," "IT AND competitive advantage," "digital transformation AND business strategy," and "IT capabilities AND firm performance" were used. No time restrictions were applied, as the review aims to trace the field's evolution from its origins in the 1960s to the present. To be included, sources had to be peer-reviewed journal articles, conference proceedings, book chapters, or authoritative institutional reports (e.g., OECD, European Commission) dealing directly with IT's role in business strategy or performance. Purely technical sources lacking a business or strategic dimension were excluded, as were non-English publications. The initial search returned over 120 potentially relevant sources, which were screened by title, abstract, and full text. The final selection comprises more than 40 sources.

The literature search used an iterative keyword strategy combining general terms with specific constructs (IT capabilities, digital transformation, competitive advantage) and Boolean operators to ensure coverage across periods. The study spans six eras, defined not only chronologically but primarily by shifts in IT’s business role, each characterized by distinct economic, organizational, and technological dimensions. Using a thematic synthesis approach, the literature is organized into these phases, within which key themes, theoretical frameworks, and empirical findings are compared to construct an integrative framework mapping IT’s strategic ascendance.

3. Research

3.1. The foundational era: IT as operational support (1960s–1980s)

The first business applications of IT focused on automating routine clerical and administrative tasks. In the 1960s and 1970s, mainframes handled payroll, inventory, and basic accounting. Choi and Chan (2000) note that IT in its early business form was seen as a support tool, making parts of the job routine more efficient. The function of producing a document stayed the same, but the process changed with new word-processing programs. What started with dot-matrix printers and simple software eventually led to networking and database resources supported by telecommunications and multimedia (Choi & Chan, 2000).

IT was widely regarded as a cost center during this period – a necessary expense for operational continuity, not a source of strategic advantage. Drucker (1963) emphasized managing for business effectiveness, laying groundwork for understanding how information could support decision-making. Parsons (1983) was among the first to argue IT could serve as a competitive weapon, a view later expanded by Ives and Learmonth (1984), who showed how information systems could be deployed strategically for market advantage.

The introduction of Transaction Processing Systems (TPS), Management Information Systems (MIS), and Decision Support Systems (DSS) marked a gradual shift from clerical automation toward managerial support. As Choi and Chan (2000) observed, while these systems initially enabled and facilitated business processes, they evolved into initiators of organizational transformation. This era laid the technological and conceptual foundations for IT's later strategic applications.

3.2. The strategic awakening: IT as competitive weapon (1980s–1990s)

The 1980s brought a paradigm shift in how organizations viewed IT. Porter and Millar (1985) argued in their work that IT transforms the value chain and creates competitive advantage, reframing it from a back-office utility to a strategic resource capable of reshaping industries. Neo (1988) explored factors facilitating IT use for competitive advantage, while Hammer and Mangurian (1987) examined the changing strategic value of communications technology.

Strategic alignment became a central theme. Henderson and Venkatraman (1993) proposed the Strategic Alignment Model (SAM), positing that performance depends on the fit between business strategy, IT strategy, organizational infrastructure, and IT infrastructure. This model became foundational for IT governance and strategic planning research. Cortes Lopez et al. (2024) note that alignment between IT and business strategy has been extensively studied, with SAM and Luftman's (2001) business-IT alignment maturity model representing classic approaches to identifying alignment determinants and their performance impact.

Yet this era also produced the productivity paradox. Despite massive IT investments in the 1980s, aggregate productivity statistics showed little improvement. Solow's (1987) famous observation crystallized the puzzle, and Brynjolfsson and Hitt (1998) worked to resolve it, demonstrating it was largely a measurement problem – traditional metrics failed to capture qualitative gains in product variety, customer service, and organizational flexibility. Brynjolfsson and Hitt (2003) later provided firm-level evidence that computing productivity was significant when measured properly, with IT investments yielding substantial returns when paired with organizational restructuring.

3.3. The internet revolution: IT as business transformation engine (1990s–2000s)

The commercialization of the Internet in the mid-1990s was a pivotal moment. It fundamentally changed the economics of information exchange, enabling new business models, disrupting traditional value chains, and creating entirely new markets. Brynjolfsson and Smith (2000) showed the Internet enabled frictionless commerce by dramatically reducing search costs and information asymmetries. Porter (2001) argued that while the Internet was powerful, it made traditional strategy more essential, not obsolete.

E-commerce transformed business operations across all functions. Lawlor (2011) noted that many earlier technologies remained in use, but the Internet changed how they were deployed, as companies expanding into global markets needed new tools for coordination. ERP systems, CRM platforms, and SCM solutions became the backbone of integrated operations. Laurindo et al. (2003) demonstrated that successful companies achieved coherence between IT investments and broader business objectives.

IT both enabled and accelerated globalization. Cairncross (1997) proclaimed "the death of distance," arguing telecommunications was eliminating geography as an economic barrier. Sassen (1998) offered a more nuanced view, examining how globalization was shaped by the new mobility of people and money through IT. Mittal and Nault (2009) showed that IT benefits extended beyond the investing firm, creating positive externalities across business networks and supply chains.

3.4. The platform era: IT as ecosystem orchestrator (2000s–2015)

In the early twenty-first century, IT evolved from an internal optimization tool to a platform for ecosystem orchestration. Kim et al. (2010) introduced IT's role in creating healthy business ecosystems, arguing that IT-facilitated symbiotic relationships among companies enable value creation no single firm could achieve alone. They identified three mechanisms: interoperability for seamless information exchange, robustness for resilience and risk management, and value creation for competitive differentiation.

Cloud computing emerged as a transformative force. The European Commission (2014) noted that cloud and industrial technologies became enablers supporting core digital forces – mobility, social business, and analytics – and emphasized that the real power lies where these forces interact. By 2014, 45% of companies had an operational SaaS (Software-as-a-Service) solution, with fully integrated cloud applications expected to become the norm (Sinhmar, 2014). Digital disruption gained prominence as digital technologies lowered entry barriers, blurred category boundaries, and empowered new entrepreneurs (European Commission, 2014). Kollmann et al. (2022) traced digital entrepreneurship's evolution from early e-commerce through Web 2.0 to the current era of blockchain, AI, and digital ecosystems.

3.5. The intelligence era: IT as strategic foundation (2015–2022)

This period saw the maturation of big data analytics, the Internet of Things, and early machine learning deployment in business. Chiunga (2025) argues that technology now plays a crucial role in transforming how businesses operate and interact with customers, enhancing efficiency, communication, and innovation. The Industrial Revolution 4.0, as Schwab (2016) described, focuses on automation and cyber-physical systems merging ICT with the industrial world (Pratama et al., 2022).

IT capabilities became central to sustainability. Lee (2023) found that IT capabilities mediate between strategic orientations and sustainable business outcomes, showing that companies with superior IT

capabilities better integrate environmental and social considerations into strategy. Business agility also received significant attention. Pahala et al. (2022) found IT essential for building enterprise capabilities to identify and respond to change – organizations with strong IT infrastructure become more responsive to market demands. Effective IT use can be the main differentiator between successful and less successful companies (Cepeda & Arias-Pérez, 2019).

Kathuria et al. (2023) revealed important nuances in how organizational forms approach IT strategy. While the positive link between IT investment and firm outcomes is well established, family-owned businesses demonstrate relatively conservative IT decision-making, investing carefully rather than aggressively. This highlighted that IT's strategic value depends not only on the technology but also on governance structures and ownership characteristics.

3.6. The autonomous intelligence era: IT as cognitive partner (2022–Present)

The release of ChatGPT in November 2022 opened a fundamentally new chapter; however, these developments should be interpreted with caution, as several forward-looking claims about autonomous and agentic AI systems remain largely speculative and are primarily reflected in industry discourse rather than established academic literature. One thing is clear – AI has shifted IT from a decision-support tool to a cognitive partner that actively participates in value creation, defined by the convergence of generative AI, hyperautomation, and agentic AI systems.

Adoption has been rapid. McKinsey's 2025 Global Survey shows 78% of organizations now use AI in at least one function (up from 55% two years prior), with 71% regularly using generative AI – most commonly in marketing, product development, service operations, and software engineering. Importantly, redesigning workflows rather than simply deploying technology drives the greatest bottom-line impact (McKinsey, 2025). PwC's 28th Annual CEO Survey (2025) reinforces this: 56% of CEOs report efficiency gains from generative AI, while about a third cite increased revenue (32%) and profitability (34%). Notably, more CEOs report growing headcount due to AI than shrinking it.

Organizations that have reached operational maturity are pulling ahead. Accenture (2024) found that "reinvention-ready" companies – with hyper-automation, AI at scale, and modern data foundations – achieve 2.5 times higher revenue growth, 2.4 times greater productivity gains, and are 3.3 times more likely to scale high-value generative AI use cases. Similarly, Bain & Company (2024) reports that companies investing at least 20% of their IT budget in automation achieve 22% cost savings versus 8% for laggards, with leaders planning nearly four times more generative AI investment.

Looking ahead, agentic AI – autonomous systems that plan, reason, and act across business processes without continuous human oversight – represents the next frontier. The global enterprise generative AI market is projected to grow from US$4.0 billion in 2025 to US$19.2 billion by 2030 (Generative AI Week, 2026), signaling IT's evolution into an autonomous cognitive layer embedded across all organizational functions.

3.7. Integrative framework: The strategic ascendance of IT

The six eras reveal a clear trajectory: IT moved from an operational cost center (Foundational Era), to a competitive weapon (Strategic Awakening), a transformation engine (Internet Revolution), an ecosystem orchestrator (Platform Era), a strategic foundation (Intelligence Era), and finally a cognitive partner (Autonomous Intelligence Era). Three dimensions consistently shape this evolution: economic, organizational and technological. Each successive era builds upon the foundations of the previous one, demonstrating that IT's strategic value is not determined by technology alone but by its interplay with economic conditions, organizational structures, and strategic alignment. Table 1 summarizes these developments across all six eras.

Table 1. Overall summary of development of IT impacting business through all six identified eras

	Economic	Organizational	Technological
Foundational Era (1960s–1980s)	IT was a pure cost center, not a strategic asset Seen as a necessary expense for operational continuity	Clerical automation through TPS, MIS, and DSS IT shifted from administrative support toward managerial decision-making Served as an initiator, facilitator, and enabler of business processes	Mainframes handled payroll, inventory, and accounting Early tools included dot-matrix printers and simple software Networking and database resources emerged over time
Strategic Awakening (1980s–1990s)	Productivity paradox emerged: massive IT spending showed little measurable improvement Traditional metrics failed to capture qualitative gains	Strategic Alignment Model (SAM) proposed: performance depends on fit between business and IT strategy IT reframed as a competitive weapon reshaping value chains	Networking and client-server architectures replaced standalone mainframes. Telecommunications integration expanded IT beyond individual departments
Internet Revolution (1990s–2000s)	Internet enabled "frictionless commerce" by reducing search costs and information asymmetries IT benefits created positive externalities across networks	ERP, CRM, and SCM became the backbone of integrated operations Successful companies aligned IT investments with broader business objectives	Internet commercialization and e-commerce platforms transformed the landscape ICT increased data transfer speed and enabled knowledge spillover effects
Platform Era (2000s–2015)	Ecosystems allowed companies to create value no single firm could achieve alone Cloud computing shifted capital expenditure to operational expenditure 45% of companies had operational SaaS by 2014	IT evolved into an ecosystem orchestrator Key mechanisms: interoperability, robustness, and value creation	Cloud, mobile, social business, and analytics converged Digital platforms lowered entry barriers and blurred category boundaries
Intelligence Era (2015–2022)	ICT capability significantly affects competitive advantage Companies with superior IT capabilities integrate sustainability into strategy and achieve stronger business outcomes	IT became essential for business agility and responsiveness Governance structures (e.g., family-owned companies) shape IT's strategic value	Big data, IoT, early ML, and Industry 4.0 merged ICT with the industrial world Business intelligence became a core competitive component
Autonomous Intelligence Era (2022–Present)	Companies with AI at scale achieve significantly higher revenue growth and productivity Strategic AI investment differentiates market leaders from laggards Enterprise AI market projected to reach US$19.2B by 2030	IT shifts from decision-support to cognitive partner Redesigning workflows drives the greatest bottom-line impact 78% of organizations now use AI in at least one function	Generative AI, hyperautomation, and agentic AI systems converge AI can instantly generate lengthy content from simple prompts Autonomous AI agents that plan, reason, and act are the next frontier

4. Discussion and conclusion

Rather than merely organizing prior literature chronologically, the six-era classification contributes theoretically by identifying recurring transformation patterns in IT's role, showing that shifts in its strategic importance are driven by interactions among economic conditions, organizational structures, and technological capabilities. The review traces IT's ascendance from a 1960s back-office cost center to an autonomous cognitive partner embedded across functions, providing both a structured synthesis and a conceptual explanation through an integrative framework.

Three key insights emerge. First, IT's economic value depends on complementary investments – the productivity paradox showed that technology alone does not generate returns without process redesign and strategic alignment (Brynjolfsson & Hitt, 1998; 2003). This remains true today, as companies redesigning workflows around AI achieve far greater impact than those merely deploying it (McKinsey, 2025). Second, organizational structures shape IT effectiveness – strategic alignment between business and IT strategy is essential (Henderson & Venkatraman, 1993), and governance characteristics such as ownership type influence IT investment patterns (Kathuria et al., 2023). Third, the technological trajectory has moved consistently toward greater convergence and autonomy, with each wave lowering barriers, expanding scope, and accelerating transformation.

If the trajectory continues, a seventh era may emerge in which IT evolves from cognitive partner to autonomous enterprise architect. AI systems would not merely support decisions but independently design, optimize, and reconfigure business models in real time. Agentic AI could evolve into multi-agent ecosystems where autonomous systems negotiate across supply chains, allocate resources, and adapt strategies without human intervention. Economically, this could produce "self-optimizing enterprises" with continuously recalibrated cost structures and pricing. Organizationally, the boundary between IT and business units may dissolve entirely. Technologically, quantum computing and advanced neural architectures could provide the foundation for enterprise-wide autonomy. However, this era would also intensify challenges around accountability, ethical AI governance, and the digital divide between organizations. As a literature review, the study is limited by source selection, and rapid AI advances may require updates. Future research should empirically test the framework across industries and examine how agentic AI reshapes human–machine decision boundaries.

References

Accenture. (2024). Reinventing enterprise operations: How reinvention-ready companies are driving growth and relevance with gen AI. Accenture Research.

Al-Badi, A., Tarhini, A., & Al-Sawaei, S. (2022). The role of business culture in the digital transformation process in the logistics industry of Oman. Journal of Logistics and Supply Chain Management, 2(1), 1–15.

Bain & Company. (2024). Automation scorecard 2024: Lessons learned can inform deployment of generative AI. Bain & Company.

Brynjolfsson, E., & Hitt, L. (1998). Beyond the productivity paradox. Communications of the ACM, 41(8), 49–55.

Brynjolfsson, E., & Hitt, L. (2003). Computing productivity: Firm-level evidence. Review of Economics and Statistics, 85(4), 793–808.

Brynjolfsson, E., Hitt, L., & Yang, S. (2002). Intangible assets: Computers and organizational capital. Brookings Papers on Economic Activity, 2002(1), 137–198.

Brynjolfsson, E., & Smith, M. (2000). Frictionless commerce? A comparison of Internet and conventional retailers. Management Science, 46(4), 563–585.

Cairncross, F. (1997). The death of distance: How the communications revolution will change our lives. Harvard Business.

Carr, N. (2003). IT doesn't matter. Harvard Business Review, 81(5), 41–49.

Cepeda, J., & Arias-Pérez, J. (2019). Information technology capabilities and organizational agility. Journal of Technology Management & Innovation, 14(2), 50–59.

Chiunga, P. T. (2025). Roles of technology in today's business environment. International Journal of Humanities Social Science and Management, 5(2), 1466–1472.

Choi, C. F., & Chan, S. L. (2000). Information technology in business: The role of IT in business processes. Business Process Management Journal, 6(3), 224–237.

Cortes Lopez, J. S., Gutierrez-Gutierrez, L., & Llorens-Montes, F. J. (2024). Information technology and business strategic alignment in small and medium organizations. Measuring Business Excellence, 28(3/4), 458–473.

Drucker, P. F. (1963). Managing for business effectiveness. Harvard Business Review, 41(3), 53–60.

Duhan, S., Levy, M., & Powell, P. (2001). Information systems strategies in knowledge-based SMEs. European Journal of Information Systems, 10(1), 25–40.

European Commission. (2014). Market analysis and foresight scenarios report. Deloitte/European Commission.

Generative AI Week. (2026). The global state of generative AI in enterprise: Industry report 2026. AgileIntel Research.

Hammer, M., & Mangurian, G. E. (1987). The changing value of communications technology. Sloan Management Review, 28(2), 65–71.

Hannan, M. (2010). The information revolution and developing countries. Georgetown University Press.

Henderson, J. C., & Venkatraman, N. (1993). Strategic alignment: Leveraging information technology for transforming organizations. IBM Systems Journal, 32(1), 4–16.

Ives, B., & Learmonth, G. P. (1984). The information system as a competitive weapon. Communications of the ACM, 27(12), 1193–1201.

Kathuria, A., Karhade, P. P., Ning, X., & Konsynski, B. R. (2023). Blood and water: Information technology investment and control in family-owned businesses. Journal of Management Information Systems, 40(1), 208–238.

Kim, H., Lee, J.-N., & Han, J. (2010). The role of IT in business ecosystem health. Communications of the ACM, 53(5), 151–156.

Kollmann, T., Hensellek, S., de Cruppe, K., & Sirber, A. (2022). Eras of digital entrepreneurship. Business & Information Systems Engineering, 64(1), 15–31.

Kraus, S., Breier, M., & Dasí-Rodríguez, S. (2020). The art of crafting a systematic literature review in entrepreneurship research. International Entrepreneurship and Management Journal, 16(3), 1023–1042.

Laurindo, F. J. B., Shimizu, T., de Carvalho, M. M., & Rabechini, R. (2003). Information technology strategy alignment: Brazilian cases. In K. Kangas (Ed.), Business strategies for information technology management (pp. 186–199). Idea Group Publishing.

Lee, V.-H., Dwivedi, Y. K., Tan, G. W.-H., Ooi, K.-B., & Wong, L.-W. (2023). How does information technology capabilities affect business sustainability? The roles of ambidextrous innovation and data-driven culture. R&D Management, 54(4), 750–774.

Luftman, J. (2001). Assessing business-IT alignment maturity. Communications of the Association for Information Systems, 4(14), 1–50.

McKinsey & Company. (2025). The state of AI: How organizations are rewiring to capture value. McKinsey Global Survey on AI.

Mittal, N., & Nault, B. R. (2009). Investments in information technology: Indirect effects and information technology intensity. Information Systems Research, 20(1), 140–154.

Neo, B. S. (1988). Factors facilitating the use of information technology for competitive advantage. Information & Management, 15(4), 191–201.

Pahala, Y., Widodo, S., Kadarusman, K., Jati, N., & Purwanto, B. (2022). The role of information technology in business agility. Quality – Access to Success, 23(189), 144–149.

Parsons, G. L. (1983). Information technology: A new competitive weapon. Sloan Management Review, 25(1), 3–14.

Porter, M. E. (2001). Strategy and the Internet. Harvard Business Review, 79(3), 63–78.

Porter, M. E., & Millar, V. (1985). How information gives you competitive advantage. Harvard Business Review, 63(4), 149–160.

Pratama, A., Meiryani, M., & Sari, M. (2022). The role of information technology in business strategy and economics in the Industrial Revolution 4.0. In Proceedings of the 2023 7th International Conference on E-Commerce, E-Business and E-Government (pp. 64–72). ACM.

PwC. (2025). PwC's 28th Annual Global CEO Survey: Reinvention on the edge of tomorrow. PricewaterhouseCoopers.

Sassen, S. (1998). Globalization and its discontents. New Press.

Schwab, K. (2016). The Fourth Industrial Revolution. World Economic Forum.

Sinhmar, V. (2014). Consequence of information technology in business. International Journal of Education & Management Studies, 4(2), 156–160.

Smith, H. A., & McKeen, J. D. (2010). How does information technology affect business value? Canadian Journal of Administrative Sciences, 10(3), 229–240.

Solow, R. (1987, July 12). We'd better watch out. New York Times Book Review, 36.

Toffler, A. (1980). The Third Wave. William Morrow.

Willcocks, L. (2010). Evaluating information technology investments: Research findings and reappraisal. Information Systems Journal, 2(4), 243–268.

RESILIENCE FOR AUTONOMOUS SYSTEMS

RE-THINKING AUTOMATION: INTENSIVE HUMAN-MACHINE COLLABORATION

Erwin Schoitsch

Center for Digital Safety & Security
AIT Austrian Institute of Technology, Vienna, Austria
erwin.schoitsch@ait.ac.at

DOI: 10.35011/IDIMT-2026-127

Keywords

Autonomous Systems; Resilience; Collaborative Safety; Artificial Intelligence; Trustworthiness; Human-Machine Teaming; Standardization.

Abstract

This paper provides an overview of an upcoming trend with respect to industrial automation: In the past, even in context of "Smart manufacturing", Artificial Intelligence and robotics, automation mainly considered the machine/plant part. Humans of course are involved, but rather as operators, in logistics, planning and control. At the shopfloor, humans had to be protected against harm caused by machines, in most cases by "fences" or other protective mechanisms. With autonomy of robots and machines, a new generation of "intelligent" mobile machines and robots allows a concept of more intensive collaboration of humans and machines on a partner/team level. This could change our attitude to AI-based systems of all kinds ("machines" are the "AI" in e.g., robotic systems, but also business, financial, healthcare, mobility or administrative systems. This changes also the perspective of functional safety, since the human becomes part of the so-called "collaborative safety", changing also the perspective of risk management. This approach is at the AI-side driven primarily by experts, e.g., in AI-Standardization (ISO/IEC JTC1 SC42 WG4), and just recently also by IEC TC65, Industrial Automation and Control standards (IEC SC65A WG24, Human-machine collaboration; IEC TC65 WG31 Industrial Agent and Multi-Agent Systems). AI driven autonomous systems raise specific ethical as well as safety and security issues. Trustworthiness is a difficult property to prove. This is also part of the Horizon Europe project RobustifAI (GA no. 101212818) which strives for development of a methodology tailored for reliable, robust and trustworthy GenAI.

1. Introduction

The rapid uptake of AI systems and autonomous decision making in Japan was further promoted by the initiative "Society 5.0", initiated by science and government in a joint effort (Schoitsch, 2019). The main targets were to fight economic stagnation in a society which for too long a time was governed by the same party and people, and at the same time fighting the consequences of an ageing society. Certain aspects of high automation, e.g., acting robots in healthcare, tagging schoolchildren by RFID devices to monitor their movements, and the like, which are accepted in the Japanese society to an extent unacceptable in Europe under our ideas of privacy ("GDPR") and self-determination, facilitated the first move. With the rapid evolution of AI technologies and their rapid take-up world-

wide, Japan is now still leading in further developing concepts towards a “suer-smart society” with a rather “science-fiction-type” vision (Schoitsch, 2025, HMCES, 2023). Ethical and societal concerns were raised by people and organizations engaged in individual freedom, privacy and human rights protection. In Europe the concerns raised were taken up by the Commission and the European Parliament by establishing the “AI-Act”, a legally binding regulation in force since August 2024, which only lacks completion of the underpinning “Harmonized Standards” requested by “Standardization Requests” from the ESOs (European Standardization Organizations, CEN, CENELEC, ETSI).

On the safety side, standards are required to support trust in safety-critical applications using AI-systems from a technical point of view (ISO/IEC JTC1 SC42, TS 22440, a joint development with IEC SC65A MT 61508, the basic functional safety standards group). IEC TC65 and their subcommittees have also taken up these challenges recently in IEC TC65 WG31 (Standard IEC 63718 Agentic AI in industrial automation) and IEC SC65A WG24 (Human-machine collaboration, Standard IEC 63662).

On the cybersecurity side, ETSI has developed standards (ETSI-SAI), but the key standards development is now in CEN/CENELEC JTC21 for the AI-Act (prEN 18282).

2. Autonomous systems and AI: Ethical recommendations

There are far reaching concepts implemented around us with autonomous systems and AI, machine learning and decision making. Conventional methods do not allow to assess requirements of functional safety, explainability and transparency of decisions taken, assessment of risks and potential impact in context of workflows in social context. Figure 1 provides some overview on several existing guidelines, recommendations and checklists.

Figure 1. Examples for standards, regulations and recommendations on AI (Update 2026-06-21)

Worthwhile to be discussed here are the documents of the high-level expert group on trustworthy AI which developed first the "Ethics Guidelines for Trustworthy Artificial Intelligence (AI)" (European Commission, 2019), and then an "Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment" (European Commission, 2020).

The Ethics Guidelines introduced the concept of Trustworthy AI, based on seven key requirements:

1. Human agency and oversight,
2. Technical robustness and safety,
3. Privacy and data governance,
4. Transparency,
5. Diversity, non-discrimination and fairness,
6. Environmental and societal well-being,
7. Accountability.

Through the Assessment List for Trustworthy AI (ALTAI), AI principles are translated into an accessible and dynamic checklist that guides developers and deployers of AI in implementing such principles in practice. ALTAI will help to ensure that users benefit from AI without being exposed to unnecessary risks by indicating a set of concrete steps for self-assessment.

3. The European AI-Act – a European regulation

The European Commission's AI Act served as a foundational regulation, outlining ten critical requirements for high-risk AI systems. These include:

1. Risk management system for AI systems [Art. 9 AI-Act],
2. Governance and quality of datasets used to build AI systems [Art. 10 Data and data governance],
3. Record keeping - built-in logging capabilities in AI systems [Art. 11 Techn. doc., Art. 12 record-keeping],
4. Transparency and information to the users of AI systems [Art. 13 Transparency and user info],
5. Human oversight of AI systems [Art. 14 Human oversight],
6. Accuracy specifications for AI systems [Art. 15 Accuracy, robustness and cybersecurity],
7. Robustness specifications for AI systems [Art. 15 Accuracy, robustness and cybersecurity],
8. Cybersecurity specifications for AI systems [Art. 15 Accuracy, robustness and cybersecurity],
9. Quality management system for providers of AI system [Art. 17 QM for providers],
10. Conformity assessment for AI systems [Art. 19 + Art. 43 Conformity Assessment].

These requirements highlight the need for harmonized standards to support compliance and ensure the safe and responsible deployment of AI technologies across Europe.

Beyond high-level ethical principles, the CEN/CENELEC JTC 21 work program translates ethics into verifiable organizational, technical, and procedural requirements across the AI lifecycle. The current work-in-progress operationalize ethics by embedding fundamental-rights considerations,

human oversight, fairness, and accountability into risk management, data governance, technical validation, and organizational processes:

- prEN 18229-1: AI trustworthiness framework – Part 1: Logging, transparency and human oversight,
- prEN 18229-2: AI trustworthiness framework – Part 2: Accuracy and robustness,
- prEN 18228: AI Risk Management,
- prEN 18281: Artificial Intelligence – Evaluation methods for accurate computer vision systems,
- prEN 18282: Artificial Intelligence – Cybersecurity specifications for AI systems,
- prEN 18283: Artificial Intelligence – Concepts, measures and requirements for managing bias in AI systems,
- prEN 18284: Artificial Intelligence – Quality and governance of datasets in AI,
- prEN 18285: AI Conformity assessment framework,
- prEN 18286: Artificial Intelligence – Quality management system for EU AI Act regulatory purposes,
- prEN 18281: Artificial Intelligence – Evaluation methods for accurate computer vision systems.

4. The RobustifAI project

Autonomous systems collaborating with humans on a team-basis take actions and decisions in front of and collaborating with humans. HCPS (Human Cyber-Physical Systems) represent the most demanding systems with respect to the robustness of GenAI due to their immediate physical impact, criticality, real-time and human interaction requirements. How can we trust the decisions, actions, behaviour, safety/security and reliability of active collaboration?

Generative Artificial Intelligence (GenAI), such as foundation models, represents a powerful and transformative class of AI capable of learning patterns from data and generating new content. However, GenAI has notable shortcomings including hallucinations or bias which can lead to misuse or hinder its widespread adoption and positive societal and economic impact.

RobustifAI [RobustifAI, 2025] aims to develop a rigorous design and deployment methodology tailored for reliable, robust, and trustworthy GenAI. The objectives targeted at are:

- Harden the GenAI system development process and the methods within to become neural-symbolic with improved performance and robustness.
- Enhance GenAI adaptability and self-monitoring ability for evolving or new environments.
- Retrofit GenAI system development process to ensure human centricity.
- Enable sustainable GenAI solutions across diverse applications.

Three key application areas are addressed by “real-world” use cases in practice:

- Autonomous Driving: Multimodal GenAI creates a comprehensive benchmark of scenarios with a humancentric approach to ensure fairness and diversity of scenarios with a safety assessment across user groups.
- Service robotics: GenAI technology is applied to a low-speed service robot developed to understand the patients and behave harmoniously with them (Elderly care home, University kitchen).

- Security Operation Center: Automation of correct and timely operations of a Security Operation Centre with human-centric view by using Large Language Models (LLM) involving threat information synthesis, intrusion detection rule generation incident analysis reporting and response plan generation.

Particular focus is on ethical issues. An Ethics Committee was installed, and regular Ethics Checks are performed and documented in separate deliverables. Ethics Check are performed on a regular basis internally and with the Project Officer and external experts.

5. Human-Machine Teaming as generic approach

Because AI capabilities ("intelligence", "autonomy") are progressing so rapidly, it is time to think about the relationship between humans and machines. One way is to analyse systematically the possible roles of humans and machines in their cooperation in a team, a human-machine team might be a unit for governance in the future (Schoitsch, 2025).

Human-Machine Teaming: who (machine or human) does what, when and how in the process.

Definition: ISO 22989: HMT is the *„…integration of human interaction with machine intelligence capabilities".* Human-Machine Teaming considers the AI-directed „intelligent" machine and the human as a team closely cooperating and achieving results by working together.

The humans as expert and the "machine" (which means "AI", maybe robot, software, etc.) build a team; the human as "user" is confronted with the TEAM, which builds a logical unit.

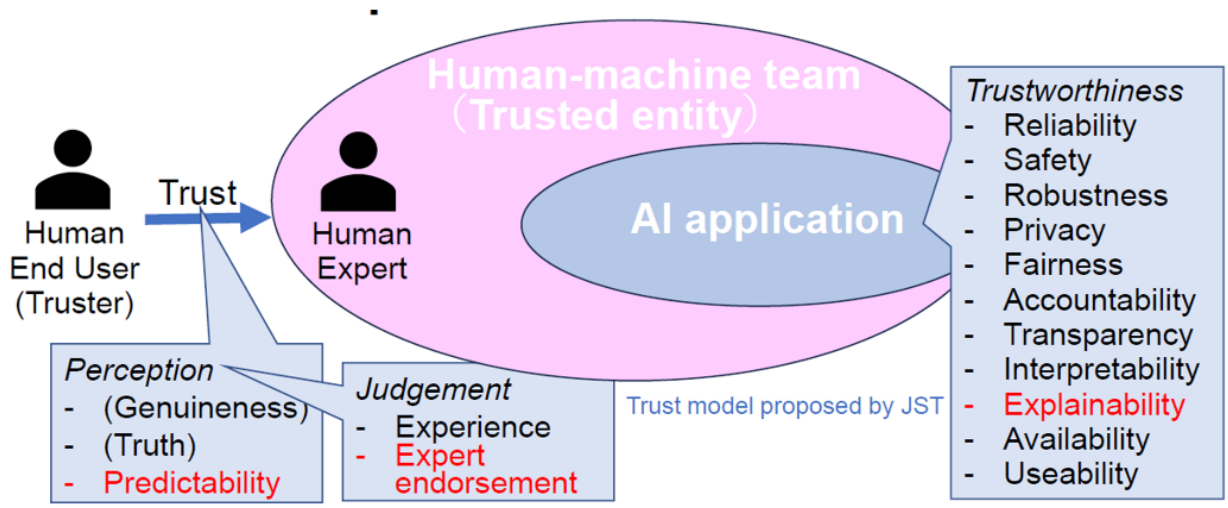


Figure 2. Human-Machine Team – Basic Configuration and impact Source: ISO/IEC JTC1 SC42 WG4 Working Drafts, Presentation at HMT WS in Brussels 2025-06-24_25 by Fumihiro Maruyama

The expert plays an important role as one partner in the team. The user is distinguished. From the users' point of view, this enhances trust of the user as "truster", the Team is more trustable because of the expert being included with respect to:

- Explainability – includes human expert,
- Predictability – through supporting human expert,
- Endorsement of the AI-system by the human expert in the team.

The distribution of work between the two partners in the team, human and machine, can be in different roles as indicated in Figure 3.

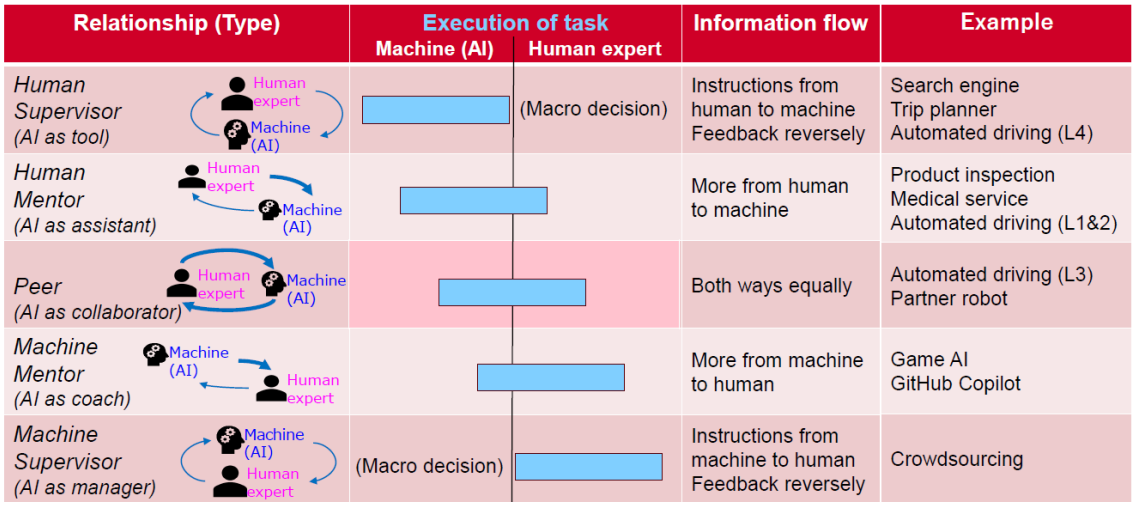


Figure 3. Relationships between Human and Machine (information flow, examples) Source: ISO/IEC JTC1 SC42 WG4 Working Drafts, Presentation at HMT WS in Brussels 2025-06-24_25 by Fumihiro Maruyama

The standards will also include technical characteristics and implementation guidelines.

6. Safety in autonomous industrial AI

"AI Everywhere?" was the motto of IDIMT 2025. AI is not a single technology of its own, it intersects with many other technology domains, as shown by the INSTAR project (Schoitsch, 2025), as presented last year at IDIMT. Therefore, many other domains have taken up AI-centred standardization activities. The joint working group between ISO/IEC JTC1 SC42 AI and IEC SC65A MT 6108 on functional safety and AI is finalizing TS 22440 (3 parts, the first one is the requirement part). The automotive (road vehicles, ISO TC22 SC32) standardization group has finished PAS 8800 (Safety and AI) and is moving towards a full standard.

IEC SC65A, Industrial measurement, control and automation, System aspects, has started 2026 a first draft of IEC 63662, Human-machine collaboration Part 1: Coordination of multiple risk assessments.

The intention of this standard is to meet the new challenges connected with new emerging technologies in ICT, namely AI enabling autonomous systems and autonomous decision taking and the resulting autonomous actions. IEC Guide 127:2026 [1] provides a guidance for developing safety standards using the concept of collaborative safety. Collaborative safety wants to achieve collaboration and coexistence between humans and machines, which helps to use effectively human capabilities, but considering the individual differences of humans in the collaboration and changing in personal condition and staff. A new approach is needed in conjunction with existing safety approaches (machinery safety and functional safety). The existing safety standards (IEC 61508, IEC 61511) consider interactions between designers and users only from the viewpoint of the system.

A key conlusion of the standard from the very beginning was (citation)

„for achieving collaboration and coexistence between humans and machines, communication between designer and user is indispensable to deal with risks dynamically changing and to utilize human capabilities. Collaborations between system elements including humans and machines during operation of the system require collaboration between stakeholders involved in the system throughout the life cycle.“

Risk assessment has to be done from all stakeholders viewpoints. This is the understanding of „multiple risk assesmnents“. The first draft bof the standard (note: all of these ideas maybe subject to changes, this is just to provide an understanding what is upcoming in this context!). Risk assesment (RA) types are defined from the stakeholders viewpoint (system includes humans and machines):

- RA-Type 1: System owner Viewpoint: interactions between stakeholders,
- RA-Type 2: System operator Viewpoint: operation of the system,
- RA-Type 3: System integrator Viewpoint: integration of the system,
- RA-Type 4: Machine supplier Viewpoint: machine in the system,
- RA-Type 5: System supplier Viewpoint: system within the SoI (System of Interest),
- RA-Type 6: Other element supplier Viewpoint: Other element,
- RA-Type 7: Operation resource manager Viewpoint: human resource.

In the draft version, RA-Typ1 to RA-Type 5 are always recommende, RA-Type 6 in case system integrator requests it, RA-Type 7 in case system operator requests it.

For each of the stakeholders, a subchapter 7.2 to 7.9 defines and interprets a set of requirements. To give an impression, how such a requirements chapter looks like (citation, system operator):

- The system operator defines the process for operating the SoI and uses the SoI for operation.
- The system operator shall conduct risk assessments in terms of operation of the SoI.
- Depending on each stakeholder, system operators should be classified into different classes, taking into account the individual differences in their abilities.
- The system operator shall take into account individual differences in operator performance in the risk assessment. A risk evaluation of the assumed range of operator competence shall be performed.
- The system operator shall consider the difference between the required competence and operator's own competence during work training. If operator's competence does not conform to the system requirements, additional safety measures shall be implemented.

Note: “shall” is a mandatory requirement, “should” a recommendation.

The draft contains a vocabulary, basic principles on concept and risk assessment, principles of modelling a system (requirements and recommendations), on information exchange between stakeholders, classification of system elements.

Another standard of IEC TC65 WG31 Industrial Agent and Multi-Agent Systems, IEC 63718 Agents in industrial automation (Agentic AI, 5 parts planned) addresses multi-agent systems in industrial context. This opens a gateway to machine-machine and machine-human collaboration on a new level. The first part “Fundamentals” focuses on modular production plants, Industry 4.0 components, hybrid energy systems and material flow and logistics as use cases.

7. Conclusions

The paper shows some important directions towards a future of industrial automation with intensive human-machine collaboration between “autonomous” intelligent systems, one the human worker, the other robots, mobile machines or other autonomous systems. It identifies ethical and societal, human

related aspects, regulations and standards underpinning a safe, trustworthy and hopefully beneficial digital workspace for us. It highlights new concepts like “Man-Machine Teaming” (HMT) which is expected to start a new area of perception and uptake of AI-driven technology (“machines”). It outlines a second development path towards collaborative safety from the industrial side. Besides the already known risks as that are addressed e.g., in the EU AI-Act, also human psychological factors and cultural bias may be a blocking factor and have to be discussed. However, we should be aware that many of the developments are upcoming anyway – we have to do our best to channel it in a proper and human-centric way.

Acknowledgements

Part of the work received funding from the EC from the EU Horizon Europe Programme, the ChipsJU Joint Undertaking and the partners’ national funding authorities (in Austria FFG (Austrian Research Promotion Agency) on behalf of BMIMI, The Federal Ministry of Innovation, Mobility and Infrastructure): AIMS5.0 (Grant agreement n° 101112089), AfarCloud (nr. 783221), A-IQ Ready (nr. 101096658), AIMS5.0 (nr. 101112089), ShapeFuture (nr. 101139996). The INSTAR Support Action is funded by Horizon Europe (nr. 101135877).

RobustifAI (Grant Agreement ID 101212818) receives funding from the European Union and the Swiss Confederation. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Health and Digital Executive Agency (HADEA). Neither the European Union nor the granting authorities can be held responsible for them. Some figures on HMT are taken from the HMT workshop presentations June 2025 at TUM in Munich, details on sources included in figure captions.

References

European Commission. (2019). High-Level Expert Group. Ethics Guidelines for Trustworthy AI. Brussels: HLEG AI. https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai

European Commission, (2020). High-Level Expert Group. Assessment List for Trustworthy AI (ALTAI). Brussels: HLEG AI. https://digital-strategy.ec.europa.eu/en/library/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment

Federal Ministry of Transport and Digital Infrastructure. Ethics Commission on “Automated and Connected Driving – Report June 2017”. Germany. https://www.bmvi.de/SharedDocs/EN/publications/report-ethics-commission-automated-and-connected-driving.pdf?__blob=publicationFile

HMCES. (2025). HCMES Guidebook, Ver.1.0.0, public release in May 2025. https://doi.org/10.18999/2012890

Japanese Government. Cabinet Office. (2018). Achieving Society 5.0. https://www8.cao.go.jp/cstp/english/society5_0/index.html

Japanese Government. (2017). Realizing Society 5.0. https://www.japan.go.jp/abenomics/_userdata/abenomics/pdf/society_5.0.pdf

Kumar, P. et al. (2022). Industry 5.0: A survey on enabling technologies and potential applications. Journal of Industrial Information Integration, 26, 100257. https://doi.org/10.1016/j.jii.2021.100257

Mayumi Fukuyama. (2018). Society 5.0: Aiming for a New Human-Centered Society. Japan Spotlight, July/Aug. 2018

RobustifAI. (2025). Robustifying generative AI through human-centric integration of neural and symbolic methods. https://robustifai.eu/

Schoitsch, E. (2019). Beyond Smart Systems – Creating a Society of the Future (5.0): Resolving Disruptive Changes and Social Challenges. In Doucek, P., Chroust, G., & Oskrdal, V. (Eds.) IDIMT 2019, Innovation and Transformation in a Digital World, Kutna Hora (pp. 387-400). Trauner Verlag, Linz, Austria.

Schoitsch, E. (2025). Human-Machine Teaming – The Way Forward to a Super-smart Society. In Doucek, P., Sonntag, M., & Nedomova, L. (Eds.) IDIMT 2025, ICT in Business: AI Everywhere? Glory and Disgrace of AI, Hradec Kralove (pp. 183-191). Trauner Verlag, Linz, Austria. https://doi.org/10.35011/IDIMT-2025-183

HEURISTIC-GUIDED DETECTION OF RESOURCE EXHAUSTION VULNERABILITIES FOR ENHANCING RESILIENCE IN AUTONOMOUS SYSTEMS

Loui Al Sardy, Mamdouh Muhammad

Computer Networks and Communication Systems (Inf7)
Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Erlangen, Germany
loui.alsardy@fau.de, mamdouh.muhammad@fau.de

Abdelkader Magdy Shaaban

Center for Digital Safety & Security
AIT Austrian Institute of Technology, Vienna, Austria
abdelkader.shaaban@ait.ac.at

Mahmoud Baklizi

Faculty of Information Technology, University of Petra, Amman, Jordan
mbaklizi@uop.edu.jo

DOI: 10.35011/IDIMT-2026-135

Keywords

Resource exhaustion; vulnerability detection; heuristic testing; system resilience.

Abstract

Autonomous and industrial control systems increasingly rely on continuous and predictable computation to maintain stable and safe operation. However, some software vulnerabilities do not cause immediate failures but instead degrade performance through excessive resource consumption. CPU-based resource exhaustion is particularly critical because it can delay decision-making processes and reduce system responsiveness, potentially affecting operational resilience. This paper presents a hybrid approach for detecting CPU-based resource exhaustion vulnerabilities. The method combines lightweight static analysis based on LLVM with heuristic-guided dynamic testing to identify input-dependent worst-case execution behaviour. Search-based optimisation techniques generate inputs that maximise resource-oriented fitness metrics, including instruction count, loop iteration frequency, and execution time. The approach was evaluated using representative algorithmic programs and a real-world utility. The results demonstrate that the framework can effectively approximate worst-case execution scenarios and reveal significant performance degradation, including slowdowns of up to 18.3×. These findings show that the proposed method can support the identification of vulnerabilities that may compromise the resilience of autonomous and industrial systems. The framework provides a practical basis for integrating resource-aware testing into existing software development and validation processes.

1. Introduction

Industrial and autonomous systems have undergone significant transformation in recent years, driven by increasing automation, connectivity, and computational capabilities. Modern Industrial Control Systems (ICS) and Industrial Automation and Control Systems (IACS) rely on interconnected software components to monitor and control physical processes in domains such as manufacturing, energy, and transportation. In such environments, reliability and stable operation are essential, as system disruptions may directly affect safety and operational continuity (Stouffer et al., 2015). With the transition towards Industry 4.0 and emerging Industry 5.0 concepts, autonomous decision-making and system intelligence have become central elements of these environments. As a result, ensuring system resilience has gained increasing importance. In this context, resilience refers not only to protection against external attacks but also to the ability to maintain acceptable performance under adverse conditions.

Distributed Denial-of-Service (DDoS) attacks are a well-known threat that disrupts services by overwhelming them with traffic (Sabottke et al., 2015; MITRE Corporation, 2020). Recent reports highlight the increasing scale and sophistication of such attacks (Kaspersky, 2022; Zayo Group, 2024). However, beyond these network-level attacks, software-level vulnerabilities can also lead to resource exhaustion. These vulnerabilities exploit inefficient execution paths within applications, causing excessive consumption of computational resources such as CPU time. Unlike traditional DDoS attacks, they do not require large traffic volumes, as carefully crafted inputs alone can trigger significant performance degradation.

In autonomous and industrial systems, such behaviour may result in delayed responses, reduced throughput, or instability in control processes. Despite their impact, these vulnerabilities are difficult to detect. Conventional approaches, including coverage-guided fuzzing and static analysis, primarily focus on crashes or memory-related faults and are less effective in identifying performance-based weaknesses (Zalewski, 2017; Deng et al., 2017). Symbolic execution techniques offer deeper analysis but are often limited in scalability.

To address these challenges, this paper proposes a hybrid analysis approach combining lightweight static analysis with heuristic-guided dynamic testing. The method leverages LLVM-based analysis to identify potentially critical code regions and applies search-based input generation guided by resource-oriented metrics such as instruction count, loop iteration frequency, and execution time.

The main contribution of this paper is a practical framework for systematically identifying CPU-based resource exhaustion vulnerabilities. While evaluated on representative programs, the approach is particularly relevant for autonomous and industrial systems, where performance degradation may directly affect system resilience and operational stability. The current work focuses specifically on CPU-oriented resource exhaustion and does not yet address memory or storage-related exhaustion behaviour.

2. Related work

The detection of resource exhaustion vulnerabilities has gained increasing attention, particularly in the context of denial-of-service attacks that exploit computational inefficiencies rather than high traffic volumes. Existing approaches range from heuristic-guided fuzzing to symbolic analysis techniques.

SlowFuzz (Deng et al., 2017) is a domain-independent fuzzing approach that guides input generation using resource-related metrics such as instruction count. This allows the identification of inefficient

execution paths without relying on crashes. However, the absence of static preselection may limit its efficiency on larger codebases. WISE (Chen et al., 2015) extends this idea by combining taint analysis with symbolic execution to prioritise inputs that lead to increased resource consumption. While this improves precision, scalability remains a challenge for complex systems. Symbolic execution tools such as KLEE (Cadar et al., 2008) enable systematic exploration of program paths and input generation based on constraints. Despite their strengths, these approaches suffer from path explosion and are not primarily designed to detect performance-related vulnerabilities. In parallel, general-purpose fuzzers such as AFL (Zalewski, 2017) and Radamsa (Oulu University Secure Programming Group, 2010) are effective in identifying crashes but rely mainly on coverage-based metrics, making them less suitable for detecting gradual performance degradation. More recent studies have examined resource exhaustion in broader contexts, including cloud and IoT systems (Kumar et al., 2024; Pietrantuono et al., 2023; Abdallah et al., 2022; Zhan et al., 2022; Du et al., 2024; Ma et al., 2024). Related work has also explored taxonomies of DDoS attacks in cloud environments (Pathak et al., 2025). These works highlight the impact on system availability but typically focus on network- or system-level detection rather than analysing the underlying software behaviour, while earlier work has explored guided testing techniques at the software level (Al Sardy et al., 2018).

In autonomous and industrial systems, such vulnerabilities are particularly critical, as performance degradation may affect control processes and decision-making. However, systematic software-level detection remains limited. The approach proposed in this paper addresses this gap by combining lightweight static analysis with heuristic-guided dynamic testing. By using LLVM-based filtering and resource-centric fitness metrics, it enables scalable identification of worst-case execution behaviour relevant to system resilience.

3. Methodology

3.1. Threat model

The proposed framework considers an adversary who is capable of providing crafted inputs to a target application with the intention of increasing its computational cost. In contrast to traditional attack scenarios, the attacker does not rely on high traffic volumes or elevated privileges. Instead, the attack exploits input-dependent behaviour within the program to trigger inefficient execution paths that result in excessive CPU consumption. Such behaviour is particularly relevant in autonomous and industrial systems, where software components are expected to operate within predictable timing constraints. If certain inputs lead to significantly increased processing time, this may affect system responsiveness and, in turn, impact control stability. In this context, excessive resource consumption refers to execution behaviour that substantially exceeds expected or average-case computational cost and may violate timing assumptions in resource-constrained environments. The focus of this work is therefore on identifying algorithmic-complexity vulnerabilities that may lead to such degradation, rather than memory-safety violations or crash-inducing faults.

3.2. General architecture

The proposed approach follows a hybrid design that combines lightweight static analysis with heuristic-guided dynamic testing. The main idea is to reduce the search space early and focus the analysis on program regions that are more likely to exhibit resource-intensive behaviour.

As illustrated in Figure , the framework consists of two main stages. The static analysis phase identifies potentially critical code regions, while the dynamic phase generates and refines inputs to explore worst-case execution scenarios. The static stage operates on the LLVM Intermediate

Representation (IR) and detects structural patterns associated with high computational cost, such as nested loops and recursion. These regions are ranked and passed to the dynamic stage, allowing the framework to concentrate on the most relevant parts of the program. The dynamic stage then applies search-based techniques to iteratively generate and evaluate inputs, with the objective of maximising resource consumption. This combination of structural filtering and guided exploration enables efficient identification of resource-intensive execution paths.

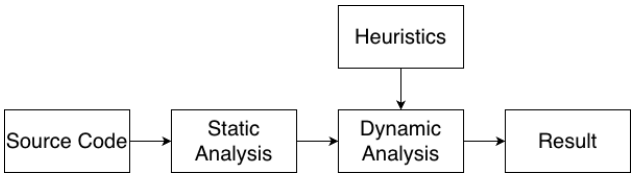


Figure 1. Hybrid workflow combining static analysis, dynamic testing, and heuristic guidance Source: authors

3.3. Static analysis using LLVM

LLVM is a modular compiler infrastructure that provides an IR suitable for program analysis across multiple compiled programming languages. The proposed approach uses LLVM IR because it offers architecture-independent structural information while remaining lightweight compared to binary-level symbolic analysis. In contrast to purely binary-based approaches, LLVM enables efficient extraction of control-flow and loop-related characteristics relevant for resource-aware testing. The static analysis component is implemented using Clang/LLVM and extracts structural information from the program under test. Rather than relying on full symbolic analysis, it uses lightweight indicators such as loop nesting depth, instruction density, and function calls within loops, which are commonly associated with increased computational complexity. An LLVM pass identifies and annotates code regions exceeding predefined thresholds, producing a set of candidate regions for further analysis. This approach reduces scalability issues while still providing effective guidance for the dynamic testing stage. The current implementation targets statically compiled programs supported by the LLVM toolchain, such as C and C++. Support for dynamically interpreted languages and binary-only software remains part of future work.

3.4. Heuristic-Guided dynamic testing

The dynamic analysis stage is responsible for generating inputs that trigger high computational cost in the previously identified code regions. The process follows an iterative structure in which inputs are evaluated and refined over multiple generations. At the beginning of the process, inputs are classified into simple categories such as numeric, string, or boolean, based on how they are used within the program. This step helps ensure that generated inputs remain syntactically valid and increases the likelihood of reaching deeper execution paths. Each input is then executed, and its impact on resource consumption is measured. A fitness value is assigned based on predefined metrics, and inputs with higher fitness are selected for further mutation and recombination. This process is similar to search-based software testing and allows the framework to gradually move towards worst-case execution behaviour. Three heuristic metrics are used to guide this process:

- *Instruction count:* which reflects the total number of executed LLVM IR instructions.
- *Loop iterations*: capturing the depth and frequency of repeated execution.
- *Execution time*: representing the overall computational effort.

These metrics provide complementary views of resource consumption and serve as optimisation signals during the search process. Inputs that result in higher values are considered more critical and are prioritised in subsequent iterations. The overall workflow consists of repeated cycles of input

generation, execution, evaluation, and mutation as shown in Figure 2. Over time, this allows the framework to converge towards inputs that expose inefficient execution paths and potential resource exhaustion vulnerabilities.

4. Evaluation

This section evaluates the effectiveness of the proposed framework in identifying resource-intensive execution paths. The analysis is guided by two main questions:

- Whether heuristic-guided testing improves the detection of high-cost execution behaviour compared to unguided approaches, and
- Fow different fitness metrics perform across varying program structures.

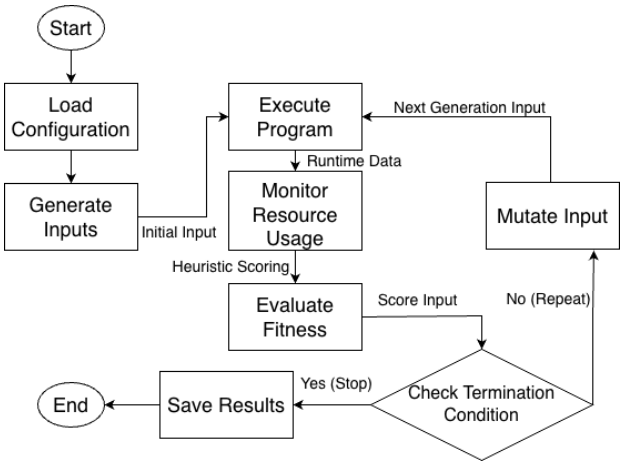


Figure 2. Iterative loop for heuristic-guided dynamic testing Source: authors

4.1. Experimental setup

The evaluation was conducted on three representative programs exhibiting input-dependent computational behaviour, including iterative, recursive, and real-world execution patterns. BubbleSort and QuickSort were selected because they represent well-understood iterative and recursive execution behaviours commonly used in complexity-oriented testing research, including prior work such as SlowFuzz. The GNU *which* utility was included to provide a lightweight real-world example exhibiting input-dependent filesystem traversal behaviour. All experiments were performed in a virtualised environment with 8 GB RAM and two CPU cores (AMD Ryzen 5 3600), running Ubuntu 20.04.4 LTS. The programs were compiled using Clang with LLVM 15.0.0 to enable the measurement of instruction counts, loop iterations, and execution time. To ensure comparability, both the genetic algorithm (GA) and particle swarm optimisation (PSO) were executed using fixed parameter settings across all experiments. These include population size, mutation rates, survival probabilities, and optimisation coefficients.

4.2. Results

The evaluation focuses on two algorithmic benchmarks (BubbleSort and QuickSort) and one real-world utility (GNU *which*). The primary metric used is the number of executed LLVM IR instructions (Metric I), as it provides a stable and interpretable measure of computational cost.

- *BubbleSort (Optimised):* The optimised implementation includes early termination, reducing execution cost for partially sorted inputs. Despite this, the framework consistently generated inputs reaching at least 93% of the worst-case instruction count across all tested input sizes. As shown in Figure 3 (a), the results range from approximately 60,000 instructions (n = 50) to 2.2 million instructions (n = 300), indicating that the approach effectively approximates worst-case behaviour even in optimised implementations.
- *QuickSort (Recursive)*: QuickSort exhibits more complex behaviour, where unfavourable pivot selection leads to deeper recursion. The results, illustrated in Figure 3 (b), show significant slowdowns compared to average-case execution. Observed slowdowns include approximately 4× at n = 50, 8× at n = 150, and up to 18.3× at n = 300. Both GA and PSO produced high-quality results, with GA showing slightly higher variability due to its mutation-driven exploration. These findings confirm that the framework can identify input patterns that induce worst-case recursion depth.
- *GNU which Utility*: To assess applicability beyond synthetic benchmarks, the GNU *which* utility was analysed. The results show that execution cost depends primarily on the number of command-line arguments rather than their specific content. Increased cost was observed only when inputs triggered deeper filesystem traversal. This behaviour indicates that the framework correctly avoids over-optimising in cases where input-dependent complexity is limited.

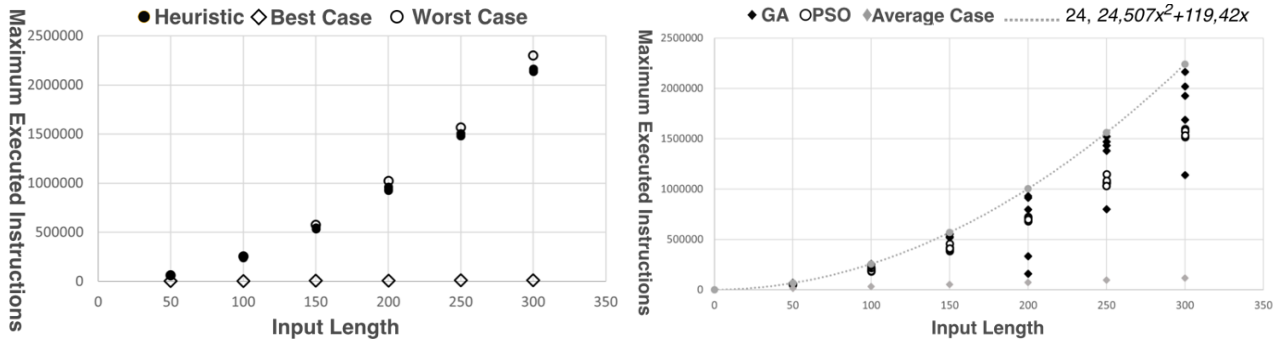


Figure 3. (a) BubbleSort and (b) QuickSort performance comparison under Metric I Source: authors

4.3. Comparison of metrics

To compare the effectiveness of the implemented fitness metrics, the optimised BubbleSort program was evaluated using instruction count, instructions within loops, and loop-execution counts. As summarised in Table 1, instruction count and instructions-in-loops show consistent and correlated growth, providing stable optimisation signals.

Table 1. Metric comparison for optimised BubbleSort using the genetic algorithm

Input Length	Executed Instructions	Instructions within Loops	Loop Execution Count	Distance Metric
50	63 013	62 805	49 441	48 712
100	147 881	245 888	191 516	123 373
150	410 720	750 977	563 019	247 042
200	960 385	1 392 739	763 890	147 702
250	1 508 196	2 056 657	1 131 198	205 978
300	2 163 894	2 174 141	1 702 570	1 673 836

Source: authors

The loop-execution metric behaves similarly but shows higher variability at larger input sizes due to early termination effects. In contrast, the distance metric produces inconsistent results and does not reflect expected complexity trends, limiting its usefulness. Overall, Metric I and Metric II provide the most reliable guidance for identifying worst-case execution behaviour.

5. Discussion

Across all benchmarks, the framework consistently generated inputs that approximate worst-case computational behaviour. Both GA and PSO proved effective, with GA providing broader exploration and PSO showing more stable convergence. Resource consumption increased only in programs with input-dependent complexity, such as BubbleSort and QuickSort, while GNU *which* remained largely unaffected. This indicates that the approach can distinguish between exploitable and non-exploitable behaviour without artificially inflating results. While runtime watchdog mechanisms may mitigate extreme delays during operation, the proposed approach aims to identify such inefficient execution behaviour earlier during development and validation. Some limitations remain. Execution time measurements may vary in noisy environments, and the current framework focuses only on CPU-related metrics. Extending the approach to additional resource dimensions, such as memory and I/O, represents a natural direction for future work.

From a resilience perspective, the results highlight how input-driven computational inefficiencies can degrade system responsiveness without causing explicit failures. In real-time autonomous systems, even moderate execution delays may become critical if they violate control-cycle timing constraints or delay sensor-processing pipelines. In autonomous and industrial systems, such delays may affect control-loop timing and decision-making processes. Identifying these behaviours during testing can therefore support more robust system design. The proposed framework provides a practical step towards integrating resource-aware analysis into resilience engineering workflows.

The proposed approach was compared with WISE (Chen et al., 2015), SlowFuzz (Deng et al., 2017), and Singularity (Wei et al., 2018). WISE is capable of generating worst-case inputs but is limited to relatively small input sizes. In contrast, the proposed framework scales to significantly larger inputs while maintaining valid results. SlowFuzz reports a slowdown of approximately 5.12× for QuickSort, which is comparable to the 4.81× slowdown observed at smaller input sizes in this work. However, the proposed approach demonstrates stronger scalability at larger inputs. Singularity achieves highly accurate worst-case approximations but requires domain-specific modelling. The proposed framework, by contrast, operates without specialised configuration, offering a more general and practical solution. Overall, the approach provides a balance between scalability, generality, and effectiveness, positioning it between symbolic methods and unguided fuzzing techniques.

6. Conclusion

This paper presented a hybrid analysis framework that combines lightweight static filtering with heuristic-guided dynamic testing to detect CPU-based resource exhaustion vulnerabilities. By leveraging LLVM-based analysis to identify critical code regions and applying search-based input generation guided by resource-oriented metrics, the approach is able to uncover worst-case execution behaviour that is typically not detected by conventional testing techniques. The evaluation demonstrates that the proposed method can reliably approximate worst-case execution scenarios across different program structures, including both algorithmic benchmarks and a real-world utility. The results show that metrics such as instruction count and loop execution provide stable guidance for identifying computationally intensive execution paths. From a broader perspective, the findings

highlight the relevance of performance-related vulnerabilities in the context of autonomous and industrial systems. Unlike traditional faults, these vulnerabilities may not lead to immediate system failure but can degrade responsiveness and affect system stability over time. Early identification of such behaviour can therefore contribute to improving system robustness. Future work will focus on extending the framework to additional resource dimensions, such as memory and I/O consumption, and exploring adaptive or learning-based heuristics. Another promising direction is the analysis of automatically generated code, for example from large language models, where inefficient execution patterns may not be immediately visible.

References

Abdallah, A. E., et al. (2022). Resource exhaustion attack detection scheme for WLAN using artificial neural network. Computers, Materials & Continua, 74(3), 5607–5623.

Al Sardy, L., Neubaum, A., & Saglietti, F. (2018). Constraint-based testing for buffer overflows. In Proceedings of SAFECOMP Workshops (LNCS), 11094, pp. 99–111. Springer.

Cadar, C., Dunbar, D., & Engler, D. (2008). KLEE: Unassisted and automatic generation of high-coverage tests for complex systems programs. In Proceedings of the 8th USENIX Conference on Operating Systems Design and Implementation (OSDI), pp. 209–224.

Chen, S., Woo, M., Brumley, D., & Song, D. (2015). Towards principled evaluation of workload-aware input selection (WISE) for resource exhaustion attacks. In Proceedings of the 24th USENIX Security Symposium, pp. 1063–1078.

Deng, D. Y., Shoshitaishvili, Y., Kruegel, C., & Vigna, G. (2017). SlowFuzz: Automated domain-independent detection of algorithmic complexity vulnerabilities. In Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security (CCS), pp. 2155–2168.

Du, Z., et al. (2024). Medusa: Unveil memory exhaustion DoS vulnerabilities in protocol implementations. In Proceedings of the ACM Web Conference (pp. 1668–1679).

Kaspersky. (2022). DDoS attacks: Trends and statistics, Q1 2022. https://securelist.com/ddos-report-q1-2022/107190

Kumar, S., Dwivedi, M., Kumar, M., & Gill, S. S. (2024). A comprehensive review of vulnerabilities and AI-enabled defense against DDoS attacks for securing cloud services. Computer Science Review, 53, 100661.

MITRE Corporation. (2020). CWE-400: Uncontrolled resource consumption (resource exhaustion). Retrieved from https://cwe.mitre.org/data/definitions/400.html

Oulu University Secure Programming Group. (2010). Radamsa fuzzer. Retrieved from https://gitlab.com/akihe/radamsa

Pathak, A. K., Dinker, A. G., & Sharma, S. K. (2025). A taxonomy of DDoS attacks on cloud computing environment and approaches to mitigation. In Proceedings of the International Conference on Cognitive Computing, Engineering, Communication, Science and Biomedical Health Informatics (IC3ECSBHI), pp. 1208–1213.

Pietrantuono, R., Ficco, M., & Palmieri, F. (2023). Survivability analysis of IoT systems under resource exhausting attacks. IEEE Transactions on Information Forensics and Security, 18, 3277–3288.

Sabottke, M., Suciu, O., & Dumitras, T. (2015). Vulnerability disclosure in the age of social media: Exploiting Twitter for predicting real-world exploits. In Proceedings of the USENIX Security Symposium.

Wei, J., Chen, J., Feng, Y., Ferles, K., & Dillig, I. (2018). Singularity: Pattern fuzzing for worst case complexity. In Proceedings of the 26th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering (ESEC/FSE), pp. 213–223.

Zalewski, M. (2017). American Fuzzy Lop (AFL). https://lcamtuf.coredump.cx/afl

Zayo Group. (2024). Protecting your business from cyberattacks – The state of DDoS attacks. https://www.zayo.com/resources/the-state-of-ddos-attacks-2024/

Zhan, M., Li, Y., Yang, H., Yu, G., Li, B., & Wang, W. (2022). CODA: Runtime detection of application-layer CPU-exhaustion DoS attacks in containers. IEEE Transactions on Services Computing, 16(3), 1686–1697.

AN INTEGRATED AI ASSURANCE AND RISK MANAGEMENT FRAMEWORK FOR AI-ENABLED TOOLS AND DIGITAL WORKSPACES IN AIMS5.0

Abdelkader Magdy Shaaban, Christoph Schmittner

AIT Austrian Institute of Technology, Vienna, Austria
abdelkader.shaaban@ait.ac.at, christoph.schmittner@ait.ac.at

DOI: 10.35011/IDIMT-2026-143

Keywords

AI Assurance; Industry 5.0; trustworthy AI; risk management.

Abstract

AI-enabled solutions based on IoT, Semantic Web ontologies, machine learning, and artificial intelligence are central to the AIMS5.0 - Artificial Intelligence in Manufacturing leading to Sustainability and Industry5.0 project's vision of a human-centric and climate-friendly transition towards Industry 5.0. In such socio-technical settings, it is no longer sufficient to assess the trustworthiness of AI components in isolation; user acceptance and operational system success are equally critical for sustainable deployment. Existing approaches, however, tend to treat these dimensions separately. This paper proposes an integrated AI assurance and risk management framework developed for AIMS5.0, which connects four established perspectives into one structured assurance logic: the Technology-Organisation-Environment (TOE) framework for context definition and AI risk identification; the Assessment List for Trustworthy Artificial Intelligence (ALTAI) for translating risks into governed controls; the Unified Theory of Acceptance and Use of Technology (UTAUT) for user-acceptance evaluation; and the DeLone and McLean model for system-success evaluation. A risk-to-control and evaluation matrix operationalises these layers by linking each identified risk to the relevant ALTAI requirement, the corresponding control, the applicable UTAUT and DeLone and McLean dimensions, and the expected assurance outcome. The contribution is conceptual and methodological: it provides dual-branch assurance logic and an evidence-oriented mapping mechanism that supports trustworthy, accepted and successful AI in AIMS5.0 digital workspaces.

1. Introduction

The transition from Industry 4.0 to Industry 5.0 is reshaping industrial environments around human-centricity, sustainability, and resilience, while preserving the high levels of connectivity and automation introduced by earlier industrial paradigms. Industry 5.0 is a value-driven evolution of Industry 4.0 that uses research and innovation to drive a sustainable, human-centric and resilient European industry, pursuing societal goals beyond growth while respecting planetary boundaries and placing worker wellbeing at the centre of production (Breque et al., 2021). Within the AIMS5.0 EU project (AIMS5.0, 2023), this transition is operationalised through a set of digital workspaces, which represent the project's concrete use cases and application environments. These digital workspaces are

supported by AI-enabled tools that integrate Internet of Things (IoT) technologies, Semantic Web ontologies, machine learning, and artificial intelligence to improve working conditions, strengthen human-centred industrial processes, and promote climate-friendly production. The growing reliance on such AI components, however, introduces a wide spectrum of risks, ranging from technical robustness and cybersecurity to fairness, privacy, governance and societal impact.

In response, policy and standardisation initiatives have promoted trustworthy AI frameworks, including the Ethics Guidelines for Trustworthy AI (European Commission, 2019), and most notably, the Assessment List for Trustworthy Artificial Intelligence (ALTAI) (European Commission, 2020). At the same time, the information systems literature offers well-established models for evaluating technology in operational settings: The Unified Theory of Acceptance and Use of Technology (UTAUT) (Venkatesh et al., 2003) addresses likely user adoption, while the updated DeLone and McLean (D&M) Information Systems Success Model (DeLone & McLean, 2003) addresses system-level value creation. In addition, the Technology-Organisation-Environment (TOE) framework (Baker, 2011) provides a structured means of understanding project complexity by helping project teams anticipate and recognise potential complexities that may emerge during the course of a project. It is particularly valuable in the early phases of a project and should ideally be applied collaboratively and revisited over time, as project complexity is both dynamic and shaped by stakeholder perception (prod-edxapp, 2018). Although each of these perspectives is well established, they are typically applied separately. For AIMS5.0, which must simultaneously ensure trustworthy AI behaviour and demonstrate value in concrete digital workspaces, this separation is insufficient.

This paper is based on the activities reported in AIMS5.0 Deliverable D7.3, Recommendations based on the validation of the generated consolidated guidelines regarding change management to Digital Workspaces. Building on that work, this paper proposes an integrated AI assurance framework that combines a TOE-based risk identification layer with two complementary branches of analysis. The first branch addresses whether the AI tools developed within AIMS5.0 are trustworthy and governed through appropriate controls, drawing on the ALTAI framework. The second addresses whether the project's digital workspaces, understood as the relevant use cases, demonstrate user acceptance, effective use, and system success in practice, drawing on the UTAUT model and the DeLone and McLean Information Systems Success Model. An integrated risk-to-control and evaluation matrix operationalises the relationships between these layers by linking each identified risk to the relevant ALTAI requirement, the corresponding control, the applicable UTAUT and DeLone and McLean dimensions, and the expected assurance outcome. The novelty lies not in the individual frameworks themselves, but in their combination within a unified assurance logic tailored to AIMS5.0 digital workspaces and AI tools.

The four frameworks were selected because, together, they cover the four non-overlapping assurance questions raised by AI-enabled digital workspaces in AIMS5.0: where do AI risks originate (TOE), what makes the AI tool itself trustworthy (ALTAI), will users actually adopt it (UTAUT), and does the deployed system create operational value (DeLone and McLean). Other governance approaches typically address only one of these dimensions, so an integrated logic spanning all four is required for AIMS5.0.

2. Related work and research gap

The proposed framework builds on four established streams of research: technology adoption determinants (TOE), trustworthy-AI assessment (ALTAI), individual-level technology acceptance (UTAUT) and information systems success (DeLone and McLean). This section briefly reviews each stream and identifies the gap the proposed integration seeks to address.

2.1. Technology-Organisation-Environment (TOE) framework

The TOE framework is described in the literature as a classical organisation-level framework for understanding the factors that influence an organisation's adoption of new technology (Jere & Ngidi, 2020). It is structured around three contexts: technological, organisational, and environmental. Within this structure, the technological elements capture complexity related to project scope and content, the organisational elements reflect internal organisational conditions, and the environmental elements represent external factors beyond the project itself (prod-edxapp, 2018). In the present work, however, TOE is not used in its traditional role as an adoption model; instead, in line with the AIMS5.0 concept, it is repositioned as a risk-definition layer from which the AI risk landscape is derived across technological, organisational, environmental, ethical, societal, data-related, and operational dimensions.

2.2. Trustworthy AI and the ALTAI Self-Assessment

Trustworthy-AI guidance has become increasingly influential in recent years, with the Assessment List for Trustworthy Artificial Intelligence (ALTAI) (European Commission, 2020) providing one of the most widely referenced operational self-assessment tools. ALTAI translates high-level ethical principles into seven requirements that define what a Trustworthy AI system must demonstrate across its lifecycle, each supported by concrete controls, safeguards, and governance practices. ALTAI has been critically reviewed in recent literature, which recognises its value as a self-assessment baseline while pointing to notable limitations. In particular, a self-assessment tool of this kind has limited utility for establishing an acceptance of trust on its own,and does not differentiate between organisations at different levels of sophistication or across application areas (Radclyffe et al., 2023).

2.3. User Acceptance of technology and UTAUT

User acceptance has been a central theme of information systems research, as market research suggests that the success rate of new technology adoption in organisations, whereby technologies bring the expected return on investment, is below 30 percent. To offer a comprehensive view of technology acceptance, (Venkatesh et al., 2003) introduced the Unified Theory of Acceptance and Use of Technology (UTAUT), which integrates key constructs from earlier technology acceptance theories (Marikyan & Papagiannidis, 2021). UTAUT has been widely applied to study technology adoption, but it was not originally designed with AI-specific concerns in mind. In particular, it captures user-based attitudes towards using a technology, but does not, on its own, specify trust-related traits that define the trustworthiness of an AI (Stevens & Stetson, 2023).

2.4. Information systems success and the DeLone and McLean model

Information system success has been a central focus in information systems and information technology research for several decades, historically complicated by diverse perspectives and fragmented measures (Jere & Ngidi, 2020). To address this fragmentation, DeLone and McLean consolidated previous measures into a unified model in 1992 (DM1992), which conceptualised success as a multidimensional construct comprising six core dimensions: information quality (IQ), system quality (SQ), user satisfaction (US), system usage (SU), individual impact (II), and organisational impact (OI). Rather than functioning as isolated metrics, these dimensions were posited to interact through specific functional pathways, linking technical and informational characteristics to user behaviour and attitudes, and ultimately to individual and organisational outcomes (Jeyaraj, 2020).

The model was subsequently revised into DM2003 (DeLone & McLean, 2003) to reflect the evolving role of information technology. Three structural changes were introduced: intention to use (IU) was added as an alternative measure for system usage; service quality (RQ) was introduced as an additional dimension directly affecting both system usage and user satisfaction; and net benefits (NB) consolidated and replaced the earlier individual-impact and organisational-impact dimensions. Both the original and the updated models have been extensively utilised and empirically supported in subsequent research evaluating diverse technologies across various contexts.

2.5. Research gap

The four streams above offer complementary but partial perspectives. TOE supports contextual and risk-oriented reasoning; ALTAI supports trustworthy-AI controls; UTAUT and DeLone and McLean support user-acceptance and system-success evaluation. In AIMS5.0, these perspectives need to be combined because the project must address two distinct but interrelated questions:

- *Are the AI tools developed in the project trustworthy and subject to appropriate controls? and*
- *Do the project's digital workspaces demonstrate user acceptance, effective use, and system success in practice?*

No single framework among the four answers both questions. The framework presented in the following sections therefore positions TOE as a shared risk-identification basis and separates the analysis into two branches, each aligned with one of the two questions, and connects them through an integrated mapping matrix.

The two-question structure reflects the key AIMS5.0 boundary: tool-level trustworthiness versus use-case-level outcomes. Within the use-case branch, UTAUT and DeLone and McLean remain distinct: UTAUT captures acceptance as a prerequisite for adoption, while DeLone and McLean captures realised system success.

3. Theoretical foundations for the proposed approach

This section explains the role each framework plays in the proposed approach. Section 3.1 describes how TOE is proposed as a structured basis for determining which risks are relevant for each AI-enabled tool and digital workspace. Section 3.2 describes how ALTAI is used to answer the first question posed by AIMS5.0, whether the AI tools developed in the project are trustworthy and subject to appropriate controls. Sections 3.3 and 3.4 describe how UTAUT and the DeLone and McLean model are jointly used to answer the second question, whether the project's digital workspaces demonstrate user acceptance, effective use, and system success in practice.

3.1. TOE as Context-definition and Risk-identification layer

In the proposed framework, TOE does not function as a trustworthiness or adoption model by itself; rather, it functions as the project's risk-definition layer, because it offers a systematic way to examine the technical conditions, the organisational setting and the external environment that shape how risks emerge (Jere & Ngidi, 2020). Within AIMS5.0, this TOE-based phase yields a multi-dimensional risk landscape covering ethical, technical, societal, data-related, operational and organisational dimensions (prod-edxapp, 2018). Representative risks include robustness and cybersecurity risks, risks related to bias and autonomy, governance and accountability risks, privacy and data integrity risks, and operational risks such as fallback and safety-critical behaviour. The TOE risk landscape may include AI-external risks, such as organisational or environmental risks around an AI tool. These

are handled through governance controls outside ALTAI, mainly R7 accountability and use-case-level evaluation, so ALTAI applies only to risks directly related to the AI system.

3.2. ALTAI as Zrustworthiness-by-design and control layer

ALTAI translates trustworthy-AI principles into seven operational requirements: (R1) human agency and oversight; (R2) technical robustness and safety; (R3) privacy and data governance; (R4) transparency; (R5) diversity, non-discrimination, and fairness; (R6) societal and environmental well-being; and (R7) accountability. Each requirement is associated with concrete controls and safeguards such as human-in-the-loop oversight, robustness to attacks, fail-safe and fallback capabilities, Data Protection Impact Assessment (DPIA) and General Data Protection Regulation (GDPR) alignment, explainability of decisions, bias testing and fairness monitoring, stakeholder participation, and ethics-board-based auditability (European Commission, 2020). In the proposed framework, ALTAI functions as a trustworthiness-by-design layer. Its main value is that it turns the risks identified through TOE into governed, auditable, and explainable control requirements.

3.3. UTAUT as User-Acceptance layer

UTAUT is used to assess whether AIMS5.0 users perceive the AI-enabled solution as useful, manageable, socially supported and adequately backed by organisational and technical infrastructure. UTAUT is complementary to ALTAI: where ALTAI addresses the properties the AI system must exhibit to be trustworthy, UTAUT addresses the conditions under which users are likely to actually adopt and use it.

3.4. DeLone and McLean as System-Success layer

The D&M model is applied to evaluate whether an AI-enabled use case delivers outputs of sufficient quality, whether the system performs adequately, whether the support and service environment is acceptable, whether the system is utilised, whether users are satisfied, and whether the use case generates meaningful benefits. In this role, D&M complements UTAUT by extending the assessment beyond user acceptance and intended use to encompass system success in operational contexts.

4. Proposed integrated AI assurance framework

The framework contains two branches because AIMS5.0 must address two distinct but interrelated questions: whether the AI tools developed in the project are trustworthy and subject to appropriate controls, and whether the project's digital workspaces demonstrate user acceptance, effective use, and system success in practice. Assessing only, one leaves a gap between trustworthy design and effective deployment. Accordingly, as illustrated in Figure 1, originally defined in AIMS5.0 Deliverable D7.3, a TOE-based phase first identifies the AI risk landscape across technological, ethical, societal, organisational, data-related, and operational dimensions. This landscape feeds into the Risk-to-Control and Evaluation Matrix, the framework's integration mechanism, which links each risk to the relevant ALTAI requirement, the corresponding control, the UTAUT and DeLone and McLean evaluation dimensions, and the expected assurance outcome. As shown in Figure 1., the structure is layered rather than parallel: ATOE defines the AI risk landscape, while ALTAI, UTAUT, and DeLone and McLean serve as evaluation lenses. The risk-to-control matrix links each risk to relevant controls, acceptance factors, and success dimensions. The central link shows that trustworthiness controls are expected to influence user acceptance and perceived workspace quality.

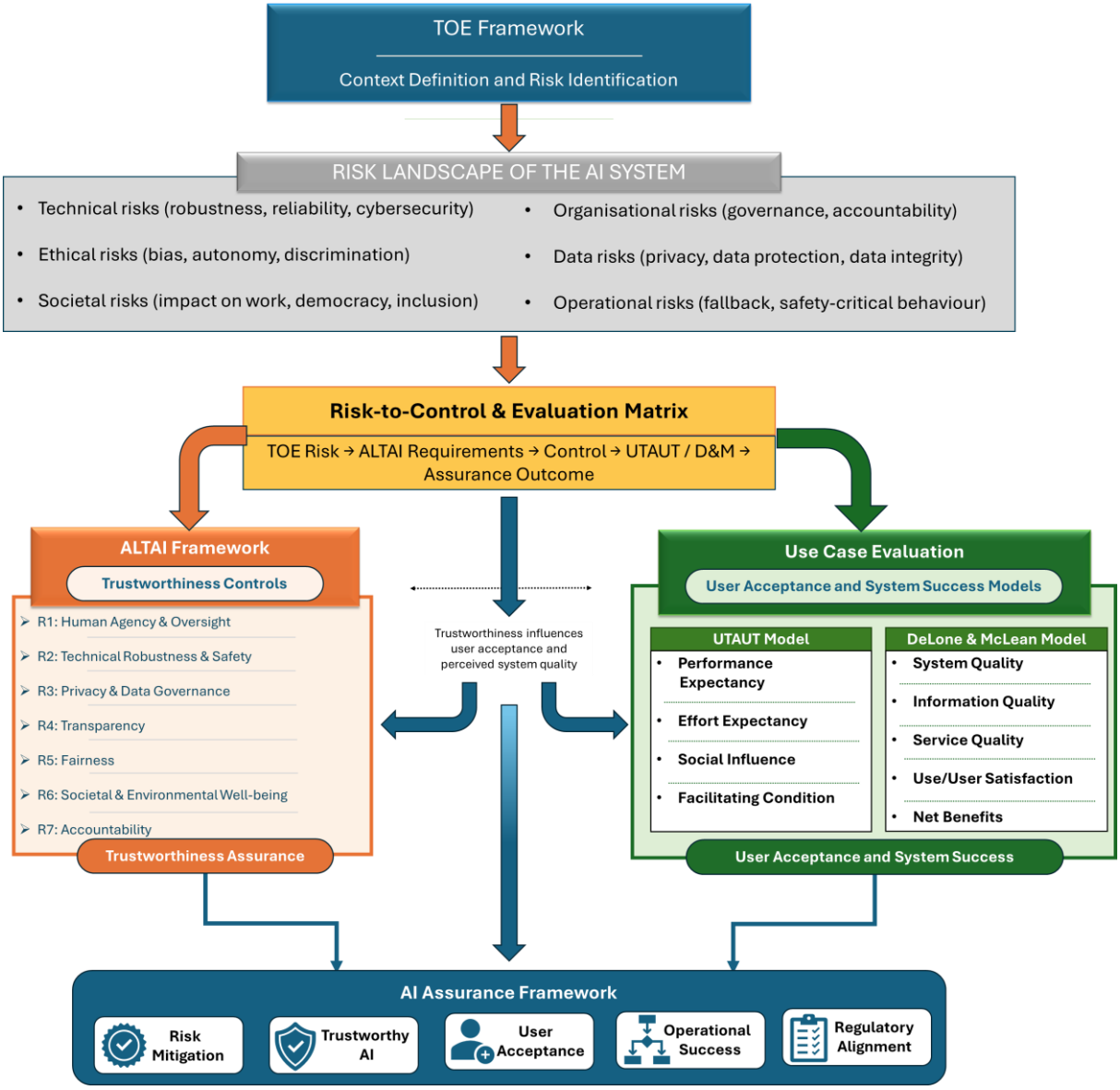


Figure 1. A high-level model of the integrated AI assurance framework for AIMS5.0 Source: authors

The model then separates into two branches. The trustworthiness branch, based on ALTAI, maps the identified risks to the relevant ALTAI requirements and controls for the AI tools. The use-case evaluation branch, based on UTAUT and the DeLone and McLean model, assesses whether the resulting solutions are accepted by users and succeed in operational environments. Both converge into one integrated AIMS5.0 AI Assurance Framework, whose intended outcomes are risk mitigation, trustworthy AI, user acceptance, operational success and regulatory alignment.

This structure is methodologically appropriate because it separates the assessment of tool-level trustworthiness controls from the evaluation of use-case-level adoption and system success, while allowing both branches to converge within a single integrated AIMS5.0 AI assurance framework. The framework also adopts the design assumption that trustworthiness influences user acceptance and perceived system quality: users are more likely to adopt and report satisfaction with AI tools that are transparent, robust, fair, and subject to accountable governance. This assumption is not presented as a universal causal claim, but as a rational integration principle underpinning the relationship between the trustworthiness and use-case evaluation branches of the proposed framework.

5. Integrated Risk-to-Control and evaluation matrix

The framework is operationalised through an integrated mapping matrix that links the risks identified for each AI system and use case to the relevant ALTAI requirements and to the corresponding UTAUT and DeLone & McLean evaluation dimensions.

Table 1. Illustrative integrated Risk-to-Control and evaluation matrix for AIMS5.0 AI-Enabled tools and digital workspaces

TOE Risk Category	Typical AIMS5.0 Risk Description	Primary ALTAI Req.	Typical ALTAI Control / Structured Solution	UTAUT Dimensions	D&M Dimensions	Main AIMS5.0 Goal
Technical	Model failure, instability, cyberattack, unreliable outputs	R2 Technical robustness and safety	Security, robustness, fallback, accuracy, and reproducibility	Effort expectancy and facilitating conditions	System quality; information quality; service quality	Reduce failures and ensure reliable AI
Ethical	Bias, autonomy, and over-reliance	R1 Oversight; R5 Fairness and non-discrimination	Human oversight, fairness, and inclusive design	Social influence; performance expectancy	User satisfaction; net benefits	Ensure human-centred and fair AI use
Societal	Work, inclusion, and societal impacts	R6 Societal and environmental wellbeing	Social review, stakeholder engagement, and safeguards	Social influence; facilitating conditions	Net benefits; user satisfaction	Prevent broader societal harm and support Industry 5.0 goals
Organisa-tional	Weak governance, unclear account-ability, poor escalation or reporting	R7 Account-ability	Governance mechanisms, auditability, ethics, and risk management	Facilitating conditions	Service quality; net benefits	Accountable deployment and evidence-based go-vernance

Source: authors

The matrix is intended as a structured decision and evidence mechanism, enabling AIMS5.0 teams to answer, for each AI tool and use case, five practical questions: (i) what is the relevant risk (TOE); (ii) which ALTAI requirement addresses it; (iii) what control or safeguard should be implemented; (iv) which acceptance or success dimensions must be checked at use-case level; and (v) what is the expected structured solution or assurance outcome. Table presents an illustrative, non-exhaustive instantiation of the matrix, as defined in AIMS5.0 Deliverable D7.3, showing the progression from risk identification through trustworthiness controls and use-case evaluation to evidence-based assurance outcomes. At the level of individual rows, the matrix provides a traceable linkage between each risk category, the corresponding controls to be implemented, and the evaluation criteria to be assessed within the relevant AI tool and digital workspace. At the overall level, it offers a structured

view of how ALTAI requirements and use-case evaluation dimensions are expected to operate across the AIMS5.0 portfolio. The matrix is intentionally concise and extensible, enabling adaptation to specific AI tools, use cases, and organisational contexts while preserving a common assurance logic.

The mapping in each row should be read as a left-to-right workflow (TOE risk → ALTAI requirement → control → use-case evaluation dimension → assurance outcome) rather than as a claim that each cell follows causally from the previous one; this row-level workflow operationalises, at the level of an individual risk, the layered structure shown in Figure 1.. The illustrative mapping is also intentionally one-to-one for readability: in practice, a given risk (such as unreliable AI outputs) typically draws on several UTAUT and DeLone and McLean dimensions simultaneously.

6. Discussion, conclusion, and future work

The proposed framework addresses a recognised gap in the assessment of AI-enabled tools and digital workspaces. Trustworthiness assessment alone, even when anchored in ALTAI, does not indicate whether users will adopt a solution or whether it will generate value in operation. Conversely, UTAUT and DeLone and McLean do not, on their own, capture the specific risk structure of AI systems or the controls required to make them trustworthy. AIMS5.0 therefore needs these perspectives brought together rather than applied in parallel.

The main contribution lies in the integration logic itself: the framework moves from risk identification to structured mitigation, assesses AI tools against an official trustworthiness reference rather than ad hoc criteria, and checks whether technically and ethically controlled systems are also accepted and beneficial in practice. In doing so, it responds to two distinct but interrelated questions: whether the AI tools are trustworthy, and whether the digital workspaces embedding them achieve user acceptance, effective use and system success, through a dual-branch assurance logic operationalised by the risk-to-control and evaluation matrix.

The framework is conceptual at this stage, so its practical value depends on empirical validation in AIMS5.0 use cases. The matrix is indicative rather than exhaustive, since real AI risks often cut across multiple categories and ALTAI requirements. The integration principle that trustworthiness influences user acceptance and perceived system quality is a reasoned design assumption rather than a proven causal law. Domain-specific tailoring is also likely to be needed, particularly for safety-critical tools, highly sensitive data, or deployments that significantly affect workforce organisation.

Future work will operationalise the framework in use cases by instantiating the matrix for at least two AI tools from different domains, with documented ALTAI controls; administering UTAUT and DeLone and McLean instruments to actual workspace users; and developing lifecycle indicators for continuous monitoring of risks, controls, acceptance, and success.

Acknowledgement

This work has received funding from the AIMS5.0 project. AIMS5.0 has been accepted for funding within the Key Digital Technologies Joint Undertaking (KDT JU), a public-private partnership in collaboration with the HORIZON Framework Programme and the national Authorities under grant agreement number 101112089.

References

AIMS5.0. (2023). AIMS5.0 - Artificial Intelligence in Manufacturing leading to Sustainability and Industry 5.0.

Baker, J. (2011). The technology–organization–environment framework. In Information Systems Theory: Explaining and Predicting Our Digital Society, 1, 231-245.

Breque, M., De Nul, L., & Petridis, A. (2021). Industry 5.0: Towards a sustainable, human-centric and resilient European industry. Publications Office of the European Union.

DeLone, W. H., & McLean, E. R. (2003). The DeLone and McLean model of information systems success: a ten-year update. Journal of management information systems, 19(4), 9-30.

European Commission. (2019). Ethics guidelines for trustworthy AI.

European Commission. (2020). Assessment list for trustworthy artificial intelligence (ALTAI) for self-assessment.

European Commission. (2022). Industry 5.0 - What this approach is focused on, how it will be achieved and how it is already being implemented.

Jere, J. N., & Ngidi, N. (2020). A technology, organisation and environment framework analysis of information and communication technology adoption by small and medium enterprises in Pietermaritzburg. South African Journal of Information Management, 22(1), 1-9.

Jeyaraj, A. (2020). DeLone & McLean models of information system success: Critical meta-review and research directions. International Journal of Information Management, 54, 102139.

Marikyan, M., & Papagiannidis, P. (2021). Unified theory of acceptance and use of technology. TheoryHub book.

Prod-edxapp. (2018). What is the TOE framework?

Radclyffe, C., Ribeiro, M., & Wortham, R. H. (2023). The assessment list for trustworthy artificial intelligence: A review and recommendations. Frontiers in artificial intelligence, 6, 1020592.

Stevens, A. F., & Stetson, P. (2023). Theory of trust and acceptance of artificial intelligence technology (TrAAIT): An instrument to assess clinician trust and acceptance of artificial intelligence. Journal of Biomedical Informatics, 148, 104550.

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view1. MIS quarterly, 27(3), 425-478.

DATA GOVERNANCE, DATA ANALYTICS & AI

FROM "GARBAGE IN → MYSTERY OUT" TO TRUSTWORTHY ANALYTICS AND AI

Data governance as an operational system of quality, context, and responsibility

Miloš Maryška, Petr Doucek, Felix Espinoza

Prague University of Economics and Business
Faculty of Informatics and Statistics
milos.maryska@vse.cz, doucek@vse.cz, felix.espinoza@vse.cz

DOI: 10.35011/IDIMT-2026-155

Keywords

Data governance; data quality; metadata management; data lineage; semantic layer (metric definitions); trustworthy AI; model governance; interpretability; auditability; reliability of data-driven decision-making; trust chain.

Abstract

The development of data analytics and artificial intelligence (AI) has increased the ability of organisations to generate predictions, recommendations, and automated decisions, but at the same time has significantly increased the demands on the quality of input data, the stability of metric meanings and the auditability of the entire decision-making chain. In practice, the phenomenon of "garbage in → mystery out" is becoming increasingly common: poor-quality, poorly contextualised, or unstable data leads to outputs that appear convincing but are difficult to explain, harder to reproduce and whose reliability is problematic for organisational management. This article proposes an integrated data governance framework that understands governance as an operating model and links three prerequisites for trust: (1) data quality as a control and evidence layer, (2) data context and semantics using metadata, lineage, and controlled metric definitions, and (3) accountability of analytical and AI artefacts through model governance, documentation, and monitoring. The result is the concept of a "trust chain" from source data to decision and a set of implementable artefacts and metrics that enable the transition from declarations to continuous recording and manage trustworthiness in a comparable way to the availability and quality of operational services.

1. Introduction

Over the past decade, data governance has shifted from a narrow concept focused on roles, access policies and cataloguing to a broader and more operational understanding: building an organisation's ability to create, maintain and securely use data and analytical outputs in decision-making. This shift is particularly clear in the context of the rapid development of AI and machine learning. While traditional business intelligence relies on aggregation and historical reporting, AI and advanced analytics are penetrating processes where the output is not just "information for humans" but a basis for decisions with a direct impact on customers, operations, or finances. This changes the risk profile: an error in the definition of a metric or a shift in upstream data may not manifest itself as "a different number in the report," but as a change in the behaviour of the model, degradation of prediction,

incorrect recommendations, or loss of stakeholder confidence. This mechanism is often associated in the literature with the fact that the greatest costs and risks do not arise in the model itself, but in dependencies and links to data, pipelines, and surrounding systems (Sculley et al., 2015).

In practice, therefore, we are increasingly seeing a phenomenon that can be summarised as "garbage in → mystery out". This formulation is deliberately sharper than the classic "garbage in → garbage out". In modern analytics, outputs are not usually "bad" at first glance; on the contrary, they can appear convincing because they are supported by sophisticated models, visualisations, or statistical terminology. The problem is that they are a "mystery": decision-makers, auditors and data teams are often unable to quickly and reliably answer questions such as: what data was used to generate the output, what metric definitions were applied, what transformations took place, whether there was any data drift or concept drift, and what are the limits and assumptions of use. In a high-stakes environment, even relatively small problems in the data can cascade downstream and accumulate in impact (Sambasivan et al., 2021). The result is usually twofold: either the organisation ignores the outputs ("we don't believe it anyway") or uses them uncritically ("the model said so"). Both strategies are risky.

The "mystery out" is usually not just the fault of the "black-box model". The ambiguity of the outputs is more often the result of a missing operating system: the organisation does not have a stabilised meaning of data, does not have continuous quality records, and does not have a controlled lifecycle of analytical artefacts. Quality is addressed reactively, and incidents cannot be quickly assessed in terms of impact; metric definitions diverge without controlled versions; and when the output starts to behave "differently," it becomes impossible to trace what has changed. This increases operational debt: every explanation of the output turns into a detective search for the origin of the data, and every change is risky because the impacts are not visible. In practice, data quality is often defined as "fit-for-purpose", i.e. whether the data is suitable for a given purpose from the user's point of view, not just from the point of view of formal correctness (Wang & Strong, 1996).

This article assumes that the credibility of analytics and AI is a property of the entire chain from data to decision, not an isolated property of a model or report. The main research question is: How can data governance be designed to systematically reduce the risk of "garbage in → mystery out" and increase the reliability of data-driven decision-making? The answer is an integrated framework that combines three capabilities:

- Data quality (DQ) as the basis for a control layer for managed correction.
- Data context and semantics through metadata, lineage, and controlled metric definitions (including a subset of the most important metrics for managing the organisation, referred to as KPIs – Key Performance Indicators).
- Responsibility for analytical and AI artefacts through model governance, documentation, and monitoring.

In the Results section, this framework is specified in operational mechanisms and two illustrative scenarios; the discussion then explains why this approach solves the "mystery out" problem systematically and what implications it has for practice and further research.

2. Methodology

The paper is conceived as design science/conceptual design (design science research), which aims to create a usable framework for reducing the risk of the "garbage in → mystery out" phenomenon in analytics and AI. The research problem is operationalised into three types of failure: (i) degradation

of data product quality, (ii) unstable semantics and context (definition of metrics, units, time anchoring), and (iii) lack of auditability and accountability of analytical and AI artefacts in the decision chain.

Based on these classes of failure, a "Chain of Trust" framework is proposed, which specifies governance mechanisms across layers from data to decision. The proposal is operationalised by a set of artefacts (a register of data products, a register of data quality rules and their records over time, a register of metrics with versioning of definitions, lineage and impact analysis, documentation of datasets and models, and change and incident processes) and measurement indicators that enable the assessment of the impact on decision-making reliability.

Two illustrative scenarios in the results serve as a demonstration of the usability of the design in different classes of problems (transactional data vs. identity/customer data) and at the same time as an ex ante evaluation in the sense of design science: operationally measurable metrics (e.g., incident rate, MTTR, reproducibility, override rate, drift/calibration) are defined for each scenario, allowing the framework to be verified in real operation without the need for a long-term study.

3. Results

The result is an integrated Trust Chain framework and its concretisation into operational and mechanisms that are capable not only of "describing" but „managing“ the trustworthiness of analytics and AI (ISO/IEC 23894:2023). The framework does not base trust on claims such as "the data is of high quality" or "the model is explainable", but on evidence: whenever an important output appears, it must be possible to answer where it came from, what exactly it means, what definitions and transformations were used, what the current state of the quality of the inputs is, and what the limits of use are. This transforms "trustworthiness" from a feeling into an operational property. The approach is compatible with how reliability is managed in technical systems: monitoring, logging, incident management, impact analysis, and change management.

The chain of trust as a problem structure

The trust chain defines three interrelated layers: the data base, the analytical/reporting layer, and the decision-making/action layer. These layers correspond to the gradual "increase in value" of data: from collection and transformation through interpretation and modelling to decision-making and action. The framework insists that trustworthiness is an emergent property: failure in one layer will very often manifest itself in another.

Data products (dataset/table/API) are created in the database, and the main output of governance is the quality control and evidence layer. In the analytical layer, data is converted into interpretation (KPIs, reports, features), and the key output of governance is stable metrics semantics and traceability (metadata, lineage). In the decision-making layer, the outputs are used, and governance here requires accountability, auditability, and a controlled lifecycle of analytical and AI artefacts.

Two cross-cutting principles are evidence instead of declarations and context as a product. Evidence means that quality, definitions, and versions are not "agreements in people's heads" but objects with history and records. Context means that definitions, units, time anchoring, exceptions, and lineage are managed as rigorously as the data itself, because without them, interpretation, reproducibility, and traceability cannot be guaranteed.

Data quality as a control and evidence layer (control plane)

The first concrete result is the redefinition of data quality from "cleaning" to a control and evidence layer in the pipeline, analogous to checking in IT operations. In practice, this means that there is a DQ registry of rules and a record of results over time, which makes it possible to distinguish random deviations from trend degradation. DQ is not defined absolutely, but according to "fit-for-purpose", i.e. according to whether the dataset meets the quality thresholds for a specific decision or model (Wang & Strong, 1996).

The DQ rule registry is not just a list of tests, but a structured description of checks linked to data products and use cases. Each rule must have a unique identity, a description of its purpose, its place in the pipeline, threshold values (including tolerances), the method of calculating metrics, and responsibility for triage.

The framework therefore introduces "DQ profiles" for use cases: the same data product may have different thresholds for freshness, completeness, or consistency depending on whether it is used for operational processes, reporting, or model recommendations. DQ profiles correspond to quality dimensions in standard data quality models (ISO/IEC 25012:2008), but their operational use is key: thresholds trigger incidents and decision-making constraints (e.g., switching the model to safe mode).

Context, semantics, and lineage as a unified meaning infrastructure

The second key part is the "context infrastructure," which stabilises the meaning of data and metrics across the organisation. Practical experience shows that a large part of the "mystery" does not arise from data errors, but from the breakdown of semantics. Organisations often have several parallel interpretations of the same term (e.g., "active customer," "revenue," "margin," "availability," "cancellation") and these interpretations gradually diverge because they are not controlled by versions and change processes. In AI, this problem is multiplied because models can be trained on a definition that does not correspond to the definition used in management; the result is a conflict: the model "works" on its data but does not meet the expectations of decision-makers.

The chain of trust assumes that, in addition to the business glossary and technical data dictionary, there is a metrics registry (semantic layer) where key KPIs (Key Performance Indicators) are defined in versions, including time anchoring, granularity, filters, and exceptions. This is a response to a frequent source of "mystery out": parallel definitions of the same metric in BI, controlling, and model feature engineering. If the definition is not versioned and controlled by a change process, the organisation cannot prove why the metric "meant something different yesterday than it does today" or reliably assess the impact of changes.

The framework therefore introduces a specific result: in addition to a business glossary and technical data dictionary, organisations also need a metric registry (semantic layer), where key KPIs are defined in a way that is both business-friendly and technically implementable. The definition of a metric must include at least the goal, meaning and time frame (when it is created and to which period it relates), granularity, filter and exceptions, units, and link to source data. It is crucial that definitions are versioned and changes undergo change management. This solves the typical problem where an organisation is unable to explain why the "same" metric appears differently in different reports. In the resulting framework, such a situation is considered a governance incident (semantic incident) rather than "normal reality".

Lineage is understood within the framework as a condition for auditability and impact analysis. The result can be formulated as a requirement to be able to answer two questions: "where did it come from" and "what will it affect". The first question is classic data lineage (upstream): sources,

transformations, definition versions. The second is impact analysis (downstream): which reports, metrics, and models are dependent on the product in question. From an operational perspective, it is essential that lineage is not just a manual description; it must be as automated as possible by collecting metadata from the pipeline and BI layer. Manual stewardship interventions in within lineage are reserved for supplementing business context and resolving exceptions, not for maintaining the basic dependency map. (Sculley et al., 2015)

Another important result is the link to incident management. If the quality of a data product is compromised, the governance system should be able to automatically determine which key metrics, reports, and models are at risk and declare a risk status accordingly. (NIST, 2023)

Governance model as product management and conditions for interoperability

The third pillar of the results is a governance proposal for analytical and AI artefacts, especially models. The framework assumes that a model is not just "code and parameters" but a product that has a life cycle and must be managed similarly to other critical services. The result is a shift from the "model exists" state to the "model is a managed product" state. The chain of trust therefore requires a minimum set of documentation and control elements: dataset documentation (datasheet), model documentation (model card), and a record of the deployment decision.

In practical terms, the framework introduces a minimum set of mandatory artefacts: dataset documentation, model documentation, and a deployment decision record. The purpose of dataset documentation is not bureaucracy, but rather to express what the dataset is, where it comes from, what its limitations are, what risks of bias or unrepresentativeness it may carry, and for what purpose it is suitable. Model documentation formalises the purpose of the model, its metrics, limitations, risks, interpretation, and expected behaviour in operation. Finally, the deployment decision record creates a governance trail: who decided, why, under what conditions, with what thresholds, and with what monitoring and retraining plan.

This approach is consistent with recommendations for dataset documentation (Gebru et al., 2018) and model property reporting (Mitchell et al., 2019). Interpretability is understood as important but conditional: explanatory techniques such as LIME or SHAP can help understand the local or global influence of variables (Ribeiro et al., 2016; Lundberg & Lee, 2017), and counterfactual explanations may be useful for communicating decision boundaries (Wachter et al., 2017), but these explanations only make sense if there is a stable context, version of definitions, and data traceability. Otherwise, the output can be "explained," but it is not certain that the explanation is reproducible and relates to the same meaning of the data.

Minimum viable governance and operational rhythm

Because governance programmes often fail due to excessive ambition, the framework provides a practical result in the form of minimum viable governance (MVG). MVG is not "simplified governance", but governance focused on enabling the organisation to quickly gain the ability to measure and manage trust in the most important data products and decisions. The resulting proposal is based on selecting a limited number of critical domains and products for which quality records, stabilised metric definitions and basic lineage with impact analysis will be introduced. This will create a functional "core" of the trust chain that can be expanded.

A significant benefit of MVG is its emphasis on a regular management rhythm. Governance without clearly defined regularity is reactive and therefore costly. The framework therefore requires a stable operating mode: systematic evaluation of data quality and trends, regular approval of changes to metric definitions, and ongoing evaluation of model status (drift, incidents, retraining plan). This

management method is scalable — operational issues are addressed on an ongoing basis, while changes to definitions and strategic decisions take place in predefined cycles. The result is governance that is active, sustainable and does not hinder development.

Measuring trustworthiness: from DQ score to "trust dashboard"

A significant result of the framework is the design of a measurement system that replaces the traditional "single DQ score" with a multidimensional view. The framework shows that trustworthiness cannot be managed with a single number, as decision-making risk arises from a combination of quality, context, and responsibility. The result is a "trust dashboard" design that typically combines three groups of indicators.

The first group describes the stability of data products: timeliness, completeness of key attributes for a given use case, consistency across systems, and anomaly trends. The second group describes reproducibility and auditability: the ability to trace the lineage of key metrics, the version status of metric definitions, and the ability to perform impact analysis in the event of an incident. The third group describes the stability of analytical and model products: drift, calibration, performance decline relative to baseline, deployment incidents, and the rate of manual override decisions. When these groups are monitored together, the organisation gains a tool that allows it to stop addressing trustworthiness subjectively and start managing it objectively as an operational attribute.

The framework also formulates an important effect: "mystery out" often arises not from a lack of data, but from a lack of traceability and versioning. Therefore, the outcome indicators also include the degree of reproducibility: how often it is possible to trace the key output (report or model score) back to the data and definitions from which it originated, and whether it is possible to recreate the output. Where reproducibility is lacking, systemic risk arises – even if the data quality looks good.

Scenario 1: Incident in transaction data → semantic conflict → drift in decision-making

The first scenario (typical for forecasting/replenishment or similar optimisation use cases) shows how a minor change in data triggers a cascade of related problems. The organisation has stable BI reporting for sales, availability, and inventory, and operates a pilot model for order recommendations in "human-in-the-loop" mode. After a period of good alignment with practice, planners notice that the model is "overshooting" for some items and "undershooting" for others, leading to an increase in manual overrides and a decline in confidence. At the same time, a discrepancy appears in the dashboards: the availability KPI is improving, while the excess inventory KPI is growing beyond expectations. Without governance, the situation is typically addressed in parallel: the BI team verifies reports, data engineering looks for errors in the pipeline, and the data science team considers retraining. The problem is that without evidence of quality, version definitions, and lineage, it quickly becomes impossible to figure out what changed first and what is merely a consequence.

The trust chain changes the process by at once framing the incident as a "decision-making incident" with roots in quality, semantics, or the model.

The first step is not retraining but looking at the trust dashboard for the model's input data products. The dashboard will show that there has been a sudden change in the distribution of values for the transactional sales data product, and at the same time, consistency between sources has deteriorated. At the same time, it will show that the completeness of the attribute (e.g., substitution group) for master data items has fallen below the threshold defined in the DQ profile for forecasting.

The second step is impact analysis: lineage will show which metrics and reports are affected by the incident and at the same time reveal a semantic conflict – the dataset for the model implicitly includes a type of transaction that is not included in the definition of "net sales" in BI. Without governance,

this would lead to a long debate about "what is correct"; with governance, it is a semantic incident: either the definitions are unified, or two metrics with a clear purpose and name are explicitly kept, and both are versioned.

The third step is risk mitigation. The governance model includes "kill switch" conditions: if the input data does not meet the thresholds (e.g., completeness of a critical attribute or detected drift), the model switches to a safer mode (e.g., conservative baseline or model without a risky feature). The fourth step is RCA supported by records. Versioning transformations and definitions will allow you to trace that on a certain date there was a change in the upstream process (e.g., accounting for stock/credit notes) that was reflected in the dataset for the model, not in the KPI definition in BI. In such a situation, retraining without correcting the semantics is risky: retraining would "cement" an inconsistent definition into the model and the incident would return in a different form (Sculley et al., 2015). Only after unifying definitions and adjusting DQ profiles (so that the incident escalates earlier next time) does it make sense to adjust features or retrain the model. In practice, governance shortens the period of "uncertainty" and reduces the cost of the incident by quickly providing answers to the questions "what has changed", "what has been affected" and "how to proceed safely".

The functionality of the Chain of Trust can be evaluated in this scenario by combining operational, analytical, and decision-making metrics. At the data base level, it is appropriate to measure trend indicators of the quality of input data products, in particular freshness relative to SLA, the rate of key quality rule breaches, and the frequency and magnitude of distribution deviations for critical variables (e.g., changes in quantiles, distribution divergence, or the proportion of "anomalous" transactions). At the semantic level, a key indicator is the number and severity of "semantics mismatch" incidents, i.e. situations where downstream outputs are based on different definitions of the same metric, and at the same time the time taken to resolve them; here it also makes sense to monitor the number of versions of definitions for key KPIs and the time from the proposal of a definition change to its approval and promotion. At the decision-making level, it is important to measure the override rate (the proportion of manual interventions in recommendations), the stability of decision-making KPIs (e.g. volatility of stocks and availability against expected seasonal patterns) and, above all, MTTR (Mean Time to Recovery) of incidents, i.e. the time from detection (alert from DQ/monitoring) to the restoration of safe use of outputs (e.g. switching to fallback mode); it is also useful to monitor the time to complete closure of the incident (root cause + prevention) separately. At the model level, drift and performance decline relative to the baseline are evaluated, as well as calibration for probabilistic outputs (Guo et al., 2017) and the "reproducibility rate", i.e. the extent to which an organisation is able to retrospectively reconstruct specific recommendations from a given version of data, definitions and model (Mitchell et al., 2019; Gebru et al., 2018). The framework can be considered functional if MTTR is demonstrably reduced over time, the number of incidents requiring manual forensic lineage tracking decreases, the reproducibility of outputs increases, and the override rate stabilises or decreases without deteriorating business KPIs. A practical criterion for "functionality" is also that incidents are detected earlier in the quality and semantics layer rather than as escalations from the business, which reduces the impact of cascades in decision-making, which the literature describes as typical for data-intensive systems (Sambasivan et al., 2021; Sculley et al., 2015).

Scenario 2: Identities and customer/account data → breakdown of segmentation → the "right" model on the wrong reality

The second scenario is deliberately contrasting: instead of transactional data and forecasting, it focuses on customer/account data and the problem of identity, deduplication, and the meaning of "customer." This is where the "mystery out" often arises differently: not as a drift over time, but as a breakdown of the entity and its semantics across systems. Organisations typically run CRM, billing, e-commerce, and loyalty programmes, with each system having its own identifiers and its own logic

of what constitutes a "customer" and what constitutes an "account". DWH or master data layers attempt to create a consolidated view (e.g., "golden customer"), but identity rules often evolve gradually, and changes are not managed as a product. As a result, entities sometimes "merge" too aggressively (false merge) or, conversely, remain divided (false split). These errors can remain hidden for a long time because reports "on aggregates" appear stable. The problem only becomes clear when an organisation starts using segmentation, personalisation or a churn model that collaborates with customer identity at a fine granularity.

Let us imagine a situation where, as part of data improvement, deduplication rules and identity resolution begin to be changed, for example, a new match key (combination of email, telephone, address) is added or the weights of the rules are changed. The goal is legitimate: to reduce the number of duplicates and improve consistency. After the change is implemented, marketing notices that the "VIP" segment has suddenly shrunk and the "new customers" segment has grown. At the same time, the churn model begins to show "better" performance on validation data, but the number of complaints and un-subscription rates increase in campaigns. Finance reports that some customers are incorrectly classified as "new" and are receiving inappropriate incentives. The teams argue: BI claims that the "number of active customers" metric is normal, marketing says that the segmentation is broken, and data science says that the model is better than before. The typical "mystery out" here is not that the numbers are wrong, but that no one has a common language for what a "customer" is and what entities are in the data.

The first step is identity management. The trust chain solves this problem by managing identities and customer/account data as a data product with explicit context and DQ profiles. The key here is the "fit-for-purpose" concept of quality: entities may be "good enough" for billing, but not for personalisation and churn modelling. Governance therefore requires that the identity product have defined service and quality parameters that are not only syntactic (NULL, format) but also semantic and statistical. A typical example is "identity stability": how often do transaction assignments to customers change over time, what is the proportion of customers with multiple accounts, what is the proportion of merges and splits, and what is the risk of false merge/false split in critical segments. These metrics are not trivial, but they can be operationalised as trend indicators and anomalies that trigger an incident. At the same time, the definition of "active customer", "new customer" and segments is formalised in the metrics registry because it is the definition of segments that bridges the gap between BI and models.

The second step is identity change. When identity resolution changes, the chain of trust requires change management and impact analysis. Lineage will show which segmentation metrics, reports, and models are directly dependent on identity. This turns identity change into a controlled release, not a "silent adjustment." In addition, such a release requires documentation: a comparison of key segments before and after, an analysis of how customer transitions between segments have changed, and how the distribution of attributes used by the model has changed. If the churn model suddenly shows better validation performance, governance asks whether this is an artefact of the change in entity definition or leakage through identifiers. This is where the importance of documenting datasets and models comes into play: the datasheet explicitly describes what the unit of analysis is (customer vs account), how identities are formed and what their limits are (Gebru et al., 2018), while the model card describes what entity definition the model applies to and what its known failure modes are (Mitchell et al., 2019). Without this anchoring, the model can be "explained" using methods such as SHAP, but we are explaining relationships on an entity that has just changed – which leads to a false sense of certainty (Lundberg & Lee, 2017).

The chain of trust also enables safe operation during an incident. If the identity proves to be unstable (e.g., a sharp increase in merging in a particular segment), marketing automation or personalisation

can be temporarily switched to conservative mode (e.g., only to a stable subset of customers, or with a limited range of actions), similar to switching the forecasting model in the first scenario. The point is not to stop business, but to reduce risk until the impact is assessed. Within the decision layer, this again supports an evidence-driven approach: the incident is not a "marketing feeling", but a state of the identity product that does not meet the DQ profile for personalisation.

This scenario shows a specific form of "garbage in → mystery out": data can be "clean" in a syntactic sense and yet unreliable for decision-making because the definition of the entity has changed. This is precisely why the chain of trust places such emphasis on context as a product and on versioning meaning. Segmentation and churn models are not just statistical; they are semantic. If the meaning of "customer" changes, so does the meaning of the KPI and the meaning of the prediction. Governance acts as a translator between worlds: it keeps stable definitions, versions, and impacts so that the organisation can make changes safely, rather than in a "wait and see" style.

The third step is evaluation. In a contrasting scenario, it is proper to base the evaluation primarily on entity stability metrics and the impact on segmentation and targeting, because the problem here does not arise as a "bad value" but as a change in the meaning of the unit of analysis.

At the product identity level, it is crucial to monitor "identity stability" over time: the rate of merging and splitting entities (merge/split rate), the proportion of customers whose transaction assignments change between periods, and the rate of increase in duplicates or, conversely, risky merging, which can be approximated, for example, by an increase in the number of entities with unusually heterogeneous attributes (e.g., multiple incompatible addresses, diverse demographic combinations) or a jump in the size of "golden record" clusters.

At the level of segmentation semantics, it is appropriate to measure segment drift: how many customers move between segments after the release of identity rules, how the distribution of key segmentation attributes changes, and whether the definition of an "active" or "new" customer changes; here, the governance metric of the speed and quality of change management of definitions (approval time, number of exceptions, number of definition conflicts) is also important.

At the model level, in addition to drift and performance, "consistency with business semantics" is evaluated: whether the model inputs and targets correspond to the same definition of entities and metrics used in reporting. from the perspective of " " practice, this will improve the stability of model explanations over time and reduce the number of situations where the interpretation of the model is based on an entity that has just been redefined (Lundberg & Lee, 2017; Ribeiro et al., 2016).

At the decision and impact level, it is relevant to monitor the rate of mis-targeting (e.g., proxy metrics such as an increase in complaints, unsubscribes, or a deterioration in conversion for campaigns targeting "new" customers), as well as an increase in manual segmentation corrections and, last but not least, the time it takes for an organisation to perform an impact analysis on reports and models after a rule identity change.

The framework can be considered functional if identity changes (and related segment changes) take place as controlled releases with predefined validation tests and "stop" thresholds, if the reproducibility of what "customer" means for a specific analytical use (datasheet and model card as standard) increases, and if the number of incidents where a "correctly" functioning model leads to inappropriate decisions due to entity change, which is a typical source of hidden operational debt (Gebru et al., 2018; Mitchell et al., 2019; Sculley et al., 2015). An important practical sign of success is that disputes between BI, marketing, and data science are shifting from subjective "what is true" to objective evidence of definition versions and measurable impacts of identity changes on segmentation.

Summary of results and measurable outputs

In summary, the framework provides three types of measurable outputs. First, the database creates trend evidence of data product quality (including "fitness for purpose") according to quality dimensions and thresholds (ISO/IEC 25012:2008). The validity of (ISO/IEC 25012:2008) was re-confirmed by ISO in 2025.

Secondly, stable semantics are created in the analytical layer: versioned definitions of metrics and traceability of their implementation via lineage, which enables impact analysis in case of an incident.

Thirdly, a governance model appears in the decision-making layer: documentation of datasets and models, records of deployment decisions and operational monitoring, including drift and calibration where probabilities are involved (Guo et al., 2017).

All three components together enable the creation of a "trust dashboard" that links the quality of inputs, the consistency of definitions, and the stability of decision outputs. This dashboard is important not only technically but also organisationally: it allows management to manage trust as an operational feature rather than a dispute over interpretation.

4. Discussion

The proposed framework can be understood as a response to a common misconception: that "governance" is primarily a matter of tools, catalogues, or compliance documents. The trust chain shows that the trustworthiness of analytics and AI only arises when quality, context, and accountability are linked to an operational system that is measurable and enforceable.

The expanded concretisation of the results shows that "garbage in → mystery out" is not primarily a technical error or an inevitable consequence of complex models, but a symptom of a missing operational system. The chain of trust offers a way to build this system: transform quality into an evidence-driven control layer, transform meaning into a versioned product, and transform models into controlled artefacts with clear conditions of use. Both scenarios also show that "mystery out" can arise through two different mechanisms. In the first scenario, it is a cascade from transactional data through semantic conflict to decision drift; in the second scenario, it is the breakdown of entity identity and thus a change in the meaning of metrics and predictions. What they have in common is that without evidence, versions, and impact analysis, the solution turns into a costly investigation and often leads to incorrect interventions (typically retraining without correcting the meaning), which is consistent with findings on the technical debt of ML systems (Sculley et al., 2015) and with the empirical picture of cascading data problems (Sambasivan et al., 2021).

The discussion also suggests that interpretability in AI is often overestimated as a standalone technique. Explanatory methods such as LIME, SHAP, or counterfactuals (Ribeiro et al., 2016; Lundberg & Lee, 2017; Wachter et al., 2017) are useful, but without a stable context and auditability, they can provide a false sense of certainty. True interpretability is "systemic": it only arises when an organisation knows what the data means, where it came from, how it was transformed, and what the limits of the model are in a specific data state. Therefore, in the chain of trust, interpretability is embedded in model governance, documentation, and monitoring, which is in line with recommendations for documenting datasets and models (Gebru et al., 2018; Mitchell et al., 2019).

The limitation of the article is that it presents a design framework and illustrative scenarios, rather than a long-term empirical study with statistical evaluation. This opens up a research programme: to measure how the introduction of quality records, a metrics registry and model governance will affect MTTR (Mean Time to Recovery) of incidents, the rate of manual interventions (overrides), the

adoption of analytics, the stability of KPIs, and the operational costs of explanation and repair. In addition to MTTR, it is also advisable to check the time to permanent repair of the root cause separately. Another direction is to find the best level of formalisation: governance must be sufficiently disciplined to provide evidence, but at the same time light enough not to block development. Minimum viable governance (MVG) offers a practical path, but its parameters (how many metrics, how many rules, what cadence) are a suitable topic for empirical comparison.

Overall, the results show that data governance can be understood not as an administrative superstructure, but as an infrastructure for reliable decision-making in the era of analytics and AI. If governance is built as an operational system, AI ceases to be a multiplier of uncertainty and becomes a scalable tool for decision-making that is defensible, auditable, and stable even in an environment of change.

Future empirical validation of the proposed framework can be conducted as an operational reliability measurement: systematic monitoring of MTTR and incident frequency, reproducibility of analytical outputs (ability to reconstruct the result from the version of data, definitions and model), override rate in decision-making processes and model drift/calibration, in order to quantify whether the Chain of Trust actually reduces operational debt and the risk of cascading failures in decision-making.

References

Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Daumé III, H., & Crawford, K. (2018). Datasheets for datasets. arXiv preprint arXiv:1803.09010.

Guo, C., Pleiss, G., Sun, Y., & Weinberger, K. Q. (2017). On calibration of modern neural networks. In Proceedings of the 34th International Conference on Machine Learning (ICML 2017).

International Organization for Standardization, & International Electrotechnical Commission. (2008). ISO/IEC 25012:2008 Systems and software engineering—Systems and software Quality Requirements and Evaluation (SQuaRE)—Data quality model. ISO.

International Organization for Standardization, & International Electrotechnical Commission. (2023). ISO/IEC 23894:2023 Information technology—Artificial intelligence—Guidance on risk management. ISO.

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting Model Predictions. In Advances in Neural Information Processing Systems.

Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I. D., & Gebru, T. (2019). Model cards for model reporting. In Proceedings of the Conference on Fairness, Accountability, and Transparency.

National Institute of Standards and Technology. (2023). Artificial Intelligence Risk Management Framework (AI RMF 1.0). U.S. Department of Commerce.

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?": Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.

Sambasivan, N., Kapania, S., Highfill, H., Akrong, D., Paritosh, P., & Aroyo, L. (2021). Data cascades in high-stakes AI. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. Everyone wants to do the model work, not the data work": Data Cascades in High-Stakes AI. Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems

Sculley, D., Holt, G., Golovin, D., Davydov, E., Phillips, T., Ebner, D., Chaudhary, V., Young, M., Crespo, J. F., & Dennison, D. (2015). Hidden technical debt in machine learning systems. In Advances in Neural Information Processing Systems.

Wachter, S., Mittelstadt, B., & Russell, C. (2017). Counterfactual explanations without opening the black box: Automated decisions and the GDPR. Harvard Journal of Law & Technology, 31(2), 841–887.

Wang, R. Y., & Strong, D. M. (1996). Beyond accuracy: What data quality means to data consumers. Journal of Management Information Systems, 12(4), 5-33.

RAG-BASED MARKET INTELLIGENCE ASSISTANT FOR BUSINESSES

A Sujith Kumar, Shalini Aggarwal, Prerna Dhawan

APEX Institute of Management, Chandigarh University, Mohali, India
24mda10075@cuchd.in, shaliniaggar@gmail.com, Erprerna_0407@rediffmail.com

Erik Boltun, Roland Weiss

Bratislava University of Economics and Business, Slovakia
Faculty of Business Economics with seat in Kosice
erik.boltun@euba.sk, roland.weiss@euba.sk

Kiran Sood

Chitkara Business School, Chitkara University, Punjab, India
Kiransood1982@gmail.com

Simon Grima

University of Malta, Faculty of Economics, Management and Accountancy, Malta
simon.grima@um.edu.mt

DOI: 10.35011/IDIMT-2026-167

Keywords

Artificial Intelligence; machine learning; natural language processing; large language models; bibliometric analysis.

Abstract

The rapid evolution of artificial intelligence has significantly transformed how organizations generate, interpret, and utilize information for strategic decision-making. Over the past decade, advancements in machine learning, natural language processing (NLP), and large language models (LLMs) have reshaped the landscape of business intelligence and market analytics. This study presents a bibliometric review of 3,378 documents indexed in the Scopus database from 1991 to 2026, aiming to map the intellectual structure, development, and collaboration trends in this domain. Using established bibliometric techniques, the research analyses publication trends, citation dynamics, keyword co-occurrence, co-citation networks, and international collaboration patterns. The findings reveal a substantial increase in research output since 2015, emphasizing the growing importance of AI-driven decision-making in business contexts. Key thematic areas include predictive analytics, decision support systems, NLP applications, and emerging generative AI. The study highlights increasing interdisciplinary collaboration and identifies key research gaps and future directions.

1. Literature review

Organizational decision-making has undergone a profound transformation over the past decades. What was once largely driven by intuition and managerial experience has increasingly become grounded in data, analytics, and computational intelligence. At the center of this shift is the growing adoption of artificial intelligence (AI) in business. From early rule-based systems to the current expansion of machine learning and large language models, AI has continuously reshaped how organizations interpret information and respond to market conditions (Chang et al., 2025). At the same time, business intelligence (BI), traditionally focused on collecting, storing, and analyzing structured data, has significantly expanded in both scope and capability. Contemporary BI systems are no longer limited to dashboards and descriptive reporting; they increasingly incorporate predictive and prescriptive functionalities supported by machine learning algorithms. As a result, organizations are moving beyond retrospective analysis toward anticipating market trends, forecasting customer behavior, and strengthening competitive positioning (Li et al., 2026). In parallel, advances in natural language processing have enabled firms to extract value from unstructured data sources such as social media content, customer reviews, and corporate reports. More recently, the rise of large language models and generative AI has accelerated this development by enabling systems to analyze, summarize, and generate insights in increasingly human-like ways (Cave, 2026; Xu et al., 2026). This is particularly relevant in market intelligence, where the rapid and accurate interpretation of large volumes of information can provide a decisive competitive advantage.

Despite the growing importance of AI-driven business intelligence, the research field remains fragmented. Existing studies are dispersed across disciplines such as computer science, management, and information systems, making it difficult to form a coherent view of the field as a whole. This fragmentation underscores the need for a systematic mapping of the available body of knowledge (Lai et al., 2026). Bibliometric analysis offers a suitable methodological response to this challenge. By examining publication patterns, citation structures, and collaboration networks, bibliometric methods make it possible to identify the development, structure, and dynamics of a research domain in a more objective and less selective way than traditional literature reviews.

This paper contributes to the literature by conducting an in-depth bibliometric analysis of research at the intersection of artificial intelligence and business intelligence (Goorts et al., 2025). Based on a dataset of 3,378 publications indexed in the Scopus database, the study examines major trends in scientific output, citation impact, thematic evolution, and scholarly collaboration. The analysis spans the period from 1991 to 2026, thereby covering both the early stages of AI integration into business environments and the more recent expansion of advanced technologies such as generative AI (Karnavas et al., 2025). The study addresses several research questions. First, how has the volume of research in this field evolved over time? Second, which themes currently dominate the literature, and which new topics are emerging? Third, who are the main contributors, and how are they connected through research collaboration? Finally, what do these developments imply for the future of AI-driven business intelligence? This study is relevant not only from a scholarly perspective but also in practical terms (Lai et al., 2026). For researchers, it provides a structured overview of the field and identifies areas requiring further investigation. For practitioners, it highlights emerging technologies and trends that may shape the future of business intelligence systems. For policymakers, it emphasizes the importance of supporting interdisciplinary research and international collaboration (Lotfipoor et al., 2026). More broadly, the evolution of AI in business intelligence reflects a wider shift toward knowledge-based decision-making. As organizations increasingly rely on intelligent systems to navigate complex and rapidly changing environments, the ability to understand and responsibly apply these technologies becomes ever more critical. This paper therefore seeks to illuminate this evolving landscape while situating AI in BI within a broader theoretical framework rooted in information

systems, decision science, and knowledge management (Mandaliya & Kulkarni, 2025; Rajan et al., 2025).

1.1. Knowledge-based view

The Knowledge-Based View (KBV) of the firm is one of the key theoretical approaches that have been used in the current research. In this kind of view, knowledge is regarded as the most strategically important resource of an organization. Knowledge is the only asset that can be expanded and developed with the use, unlike tangible assets.

The way in which data is processed, interpreted, and put in context creates knowledge that can be applied in the context of business intelligence (Seidler-de Alwis, 2026). These processes are facilitated by artificial intelligence which makes organizations to derive insights out of both structured and unstructured data and does so at a faster rate. As an example, machine learning can help to recognize the trends in customer behavior, whereas natural language processing can extract information in the form of text on social media or market reports.

In KBV terms, the AI-based business intelligence systems may be regarded as a knowledge creation and integration mechanism. They enable organizations to convert raw data into useable insights which enhances better decision making and competitive advantage. This is in accordance with the wider argument that companies that can manage and leverage knowledge successfully are in a better position to achieve in the dynamic business environment (Xu et al., 2026).

1.2. Decision support systems theory

The other theoretical basis of importance is the concept of Decision Support Systems (DSS). DSS theory focuses on the importance of information systems in supporting the decision maker with the help of relevant data, analytical tools, and models (Chang et al., 2025). DSS were traditionally considered to assist in structured decision making using predefined models with the usage of historical information. Nevertheless, their functionality has been greatly increased with the introduction of artificial intelligence. The current AI-based DSS is capable of addressing unstructured and complex issues and evolving circumstances (Mutsagondo, 2025).

The AI-based business intelligence systems are conceptualised in this research as complex versions of DSS. Not only do they give out information but also they make predictions, recommendations as well as automated decisions. This trend is a shift towards decision support to decision augmentation whereby human judgment is augmented by machine intelligence.

1.3. Resource-based view

Resource-Based View (RBV) is the complement of the KBV, as it dwells on the role of organizational resources in creating competitive advantage. The RBV states that valuable, rare, inimitable, and non-substitutable resources may give sustainable competitive advantage. Such resources can be considered artificial intelligence and data analytics capabilities. The companies that invest in modern AI technologies and train the required skills to employ AI instruments are expected to be ahead of their rivals.

2. Research methodology

2.1. Research design and approach

This research paper uses quantitative bibliometric research design to comprehensively study the transformation, structure, and the new trends in the convergence of artificial intelligence and business intelligence. The methodological tool known as bibliometric analysis has become quite accepted within the past years due to its capacity to offer a systematic and data-intensive summary of a significant mass of academic sources (Mutsagondo, 2025). As opposed to the traditional narrative reviews that are frequently based on subjective interpretation, the bibliometric approaches allow the researcher to discover patterns, relationships, and trends through objective measures, including the number of publications, citation patterns, and the frequency of the key words.

The main reason why the bibliometric approach was selected in this research is due to the fast increase in the number of studies on artificial intelligence, specifically in the fields of machine learning, natural language processing, and large language models. Such a massive and, to some extent, disjointed body of literature has a definite requirement behind having a method that could help to put everything together in a systematic manner. The bibliometric analysis makes this possible since the field can be discussed in a systematic manner with gauged indicators and this assists to give a more complete and objective view.

2.2. Data source and search strategy

The bibliographic data used in this study were retrieved from the Scopus database on [insert exact date of data collection]. Scopus was selected because it provides broad interdisciplinary coverage of peer-reviewed literature in computer science, business, management, economics, engineering and information systems, which are the main subject areas relevant to the intersection of artificial intelligence and business intelligence. In addition, Scopus offers structured bibliographic metadata, including titles, abstracts, keywords, authors, affiliations, citation data and source information, which makes it suitable for bibliometric analysis using Biblioshiny.

To ensure that the dataset captured publications at the intersection of artificial intelligence and business intelligence, the following search query was applied in the title, abstract and keywords fields:TITLE-ABS-KEY (("artificial intelligence" OR "machine learning" OR "natural language processing" OR "large language model*" OR "generative AI" OR "retrieval-augmented generation") AND ("business intelligence" OR "business analytics" OR "competitive intelligence" OR "market intelligence"))The search was not restricted at the initial stage by publication year in order to capture the historical development of the field. The final dataset covered the period from 1991 to 2026 after the filtering and cleaning process.

2.3. Collection and filtration of data

The first search produced a large number of documents; these were narrowed down by a series of filtering operations to produce quality and relevant documents. The last dataset is composed of 3,378 articles with a time frame that spans between 1991 and 2026. A number of inclusion criteria were used: Documents that were indexed by Scopus were considered only. The priority was placed on publications that were done in English to facilitate uniformity in analysis. Several types of documents were used including articles, conference papers, book chapters, and reviews. The variety of types of documents used has proven the inter-disciplinary character of the field. As an example, conference papers are especially relevant to computer science whereby new technologies are frequently described before being expanded into journal articles. Meanwhile some exclusions were made. The cleaning

process helped to eliminate documents that had an incomplete metadata or were not related to the main themes. This was needed to preserve the integrity of the dataset and also to have meaningful analysis.

3. Results and discussion

3.1. Overview of the dataset

The bibliometric analysis presented in this paper is based on a corpus of 3,378 documents retrieved from the Scopus database, covering the period from 1991 to 2026. This time frame provides a sufficiently broad perspective to capture both the early development and the recent acceleration of research at the intersection of artificial intelligence and business intelligence. The dataset indicates that the field has grown considerably in terms of both size and thematic complexity.

The annual growth rate of 13.46% should be interpreted as the average annual growth rate of scientific production over the observed period, not as an identical increase in the number of publications every year. This indicator shows that scholarly output in the field has expanded substantially, although the number of publications may vary from year to year.

The presence of 1,930 distinct sources also confirms the interdisciplinary character of the field, as the analysed publications are distributed across journals, books and conference proceedings in areas such as computer science, management, business intelligence and engineering. The dataset also includes 214,492 cited references, which indicates a broad intellectual base for the analysed research domain. At the same time, the average number of citations per document, 10.67, suggests that the publications included in the dataset have achieved visible citation activity, although this indicator should be interpreted with regard to publication year and disciplinary context.

The dominant role of conference papers (1,556) is particularly important, as it reflects the publication culture of computer science and artificial intelligence research, where conferences often serve as primary outlets for the rapid dissemination of new findings. Journal articles (935) and book chapters (446) also represent a substantial part of the dataset, indicating a balance between fast-moving technological research and more developed academic contributions.

Table 1. Overview of the dataset

Description	Results	Description	Results
Timespan	1991 - 2026	DOCUMENT TYPES	-
Sources (Journals, Books, etc)	1.930	article	935
Documents	3.378	book	170
Annual Growth Rate %	13.46	book chapter	446
Document Average Age	6.14	conference paper	1.556
Average citations per doc	10.67	conference review	165
Authors	8.634	editorial	13

Source: author's computation using Biblioshiny

3.2. Word cloud

The term cloud is used to represent the most repeated keywords in the data set and it gives a concise overview of the most prevalent ideas in the research subject. The larger words in the visualization signify the increased frequency, which means that they are central to the discourse. This number comes in handy especially when one wants to have a rough idea of the thematic focus rather than

going to an analysis level. It shows the fundamental terms of the sphere, which reflects the issues of researcher interest. But, although the word cloud is an effective way of demonstrating frequency, it lacks the ability to indicate connection between words and this weakens its analytical value. word cloud presents the distribution of the research themes. It exposes the existence of new themes, such as the natural language processing and big language models. These results indicate that the discipline is established as well as dynamic. Although some of the topics still prevail, there are always new areas of research that are constantly emerging, which is a manifestation of technological improvements. The most frequent term is "artificial intelligence" with 1,525 occurrences, confirming its central position in the analysed dataset. It is followed by "business intelligence" with 896 occurrences, "competitive intelligence" with 809 occurrences, "machine learning" with 793 occurrences and "data mining" with 538 occurrences. Other relevant terms include "learning system" with 491 occurrences and "decision support system" with 458 occurrences. These results indicate that the analysed field is strongly shaped by the interaction between AI technologies, business intelligence systems and decision-support applications.

3.3. Co-occurrence network

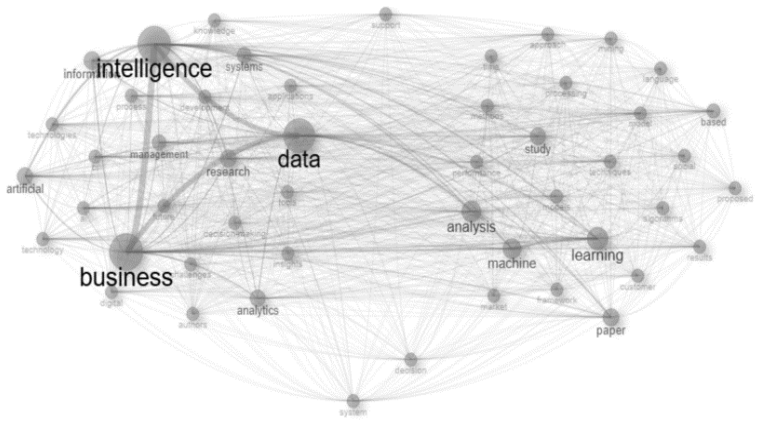


Figure 1. Co-occurrence network Source: author's computation using Biblioshiny

Figure 1 presents the co-occurrence network is a diagram which depicts the frequency of common keywords being used in same documents thus the conceptual relationships of the research field. Nodes are keywords and the relationship between them signifies the co-occurrence. The clusters of the network point to sets of the related notions, which may be related to some research topics or subdivisions. The number of connections in a cluster gives an idea of the strength of association between the keywords in the cluster, whereas intra-disciplinary connections are represented by links between clusters. This number is especially useful to know the relationship of ideas. It transcends the identification of the popular topics displaying the relationship between them and, therefore, offers the insight into how the knowledge in the field is organized. This information can assist researchers to establish dominant themes, and also the points where various topics overlap. All in all, the co-occurrence network can provide a better insight into the conceptual framework of the research field, showing both fundamental fields and emergent relationships.

The co-occurrence network gives a better insight into the relationship between various concepts. The fact that there are separate clusters in the network shows that the field exists on the basis of a few central themes. Such clusters can be related to such areas as predictive analytics, decision support systems, and market intelligence. The relations between clusters indicate the interdisciplinary character of the field where the ideas of various fields overlap and interact. This interrelatedness is the strength of this field because it enables the incorporation of different views and approaches. It

also indicates that there is a probability that further studies will be done to fill the gaps that exist among dissimilar fields.

3.4. Thematic development and evolution

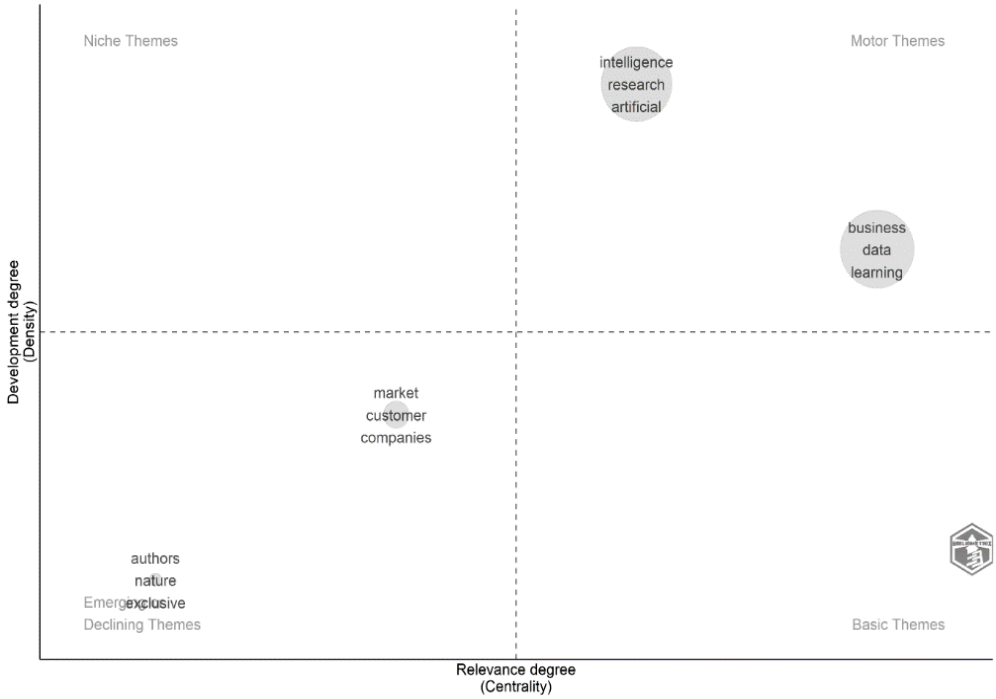


Figure 2. Thematic Map Source: Author's computation using Biblioshiny

The thematic map classifies research themes according to their centrality and density, providing insight into both their relevance and level of development. It is usually divided into four quadrants: motor themes**,** niche themes**,** emerging or declining themes**, and** basic themes. Motor themes are well-developed and central to the field, while niche themes are more specialized. Emerging or declining themes indicate areas whose relevance is changing, whereas basic themes represent important topics that may still require further development. This classification helps to understand the maturity, importance and position of individual themes within the research field. In this study, the thematic map (Figure 2) helps to show how research themes have developed over time. Established themes are mainly connected with machine learning and corporate analytics, while newer areas such as generative AI and large language models point to emerging research directions. The thematic evolution therefore reflects both continuity in established topics and innovation through new technological developments.

4. Discussion and conclusion

The findings indicate that artificial intelligence in business intelligence is a rapidly developing and increasingly influential field. Its growth has expanded the scope of research while creating new challenges and opportunities. The analysis reveals several important insights. The field continues to grow, research is becoming more interdisciplinary, and future developments are increasingly shaped by large language models and generative AI. At the same time, a balance remains between established and emerging research themes (Olubiyi et al., 2025). These findings are relevant for both academia and practice. For researchers, they highlight the importance of exploring new domains and extending existing knowledge. For practitioners, they confirm that AI-driven systems can enhance decision-

making and competitiveness (Pattabiraman et al., 2025). This paper examined the relationship between artificial intelligence and business intelligence through a bibliometric review.

The results show that this is not only a fast-growing field, but also an increasingly complex and interconnected one (Karnavas et al., 2025). The analysis points to rising scientific output, evolving research themes, and stronger collaboration, while emphasizing the transformative role of machine learning, NLP, and large language models. Future research should address the broader adoption of RAG-based systems in business intelligence, ethical implications of AI-supported decisions, real-world applications, and cross-industry comparisons (Cave, 2026). Overall, artificial intelligence is no longer a peripheral element of business intelligence, but an increasingly central one. The challenge lies not only in technological advancement, but also in its responsible use.

Acknowledgement

This research was funded by Slovak agency KEGA grant number 032EU-4/2024.

References

Cave, M. (2026). Regulating increasingly digitalised networks and services. Telecommunications Policy, 50(5), 103207.

Chang, Y. L., & Seo, B. (2025). Using explainable artificial intelligence to envision the future of work and the workforce–from prediction to explanation. Technology in Society, 103139.

Goorts, L., Hollevoet, R., Xia, V., Cammaerts, F., & Güngör, A. (2025). How do LLMs perform in the context of MCQs across different levels of thinking skills in a business education course at higher education? A comparison of ChatGPT, Gemini, and Copilot. Computers and Education: Artificial Intelligence, 100475.

Karnavas, S. I., Peteinatos, I., Kyriazis, A., & Barbounaki, S. G. (2025). Using fuzzy multi-criteria decision-making as a human-centered AI approach to adopting new technologies in maritime education in Greece. Information, 16(4).

Lai, J. Y., Wang, J., & Cai, R. H. (2026. The effects of AIoT-SC strategic alignment on SC intelligence and NPD performance. Supply Chain Forum: An International Journal, 27(1), 147-165.

Li, L., Shi, R., Zhang, R., & Chen, L. (2026). How can firms leverage responsible artificial intelligence to build competitive advantages? The role of supply chain justice and complexity. International Journal of Production Economics, 110010.

Lotfipoor, A., Fakhimi, M., Hagen-Zanker, A., & Showell, D. (2026). Large Language Models to Analyze Business Intelligence User Narratives. Journal of Computer Information Systems, 1-19.

Mandaliya, S., & Kulkarni, P. (2025). Context-Aware AI: Revolutionizing Data Engineering for Smarter Decisions. In International Conference on Computing and Machine Learning, pp. 1-17. Singapore: Springer Nature Singapore.

Mutsagondo, S. (2025). Artificial intelligence in records and information management. IGI Global.

Olubiyi, T. O., Adeyemi, O. S., Adeyemi, O. D., & Oketola, K. O. Elevating Self-Service Business Intelligence through AI and IoT in Sub-Saharan Africa: Insights from Sustainable Manufacturing Companies. In AIoT-Driven Sustainable Manufacturing and Operational Initiatives (pp. 158-173). CRC Press.

Pattabiraman, K., Sachin, A., & Lokeshkumar, R. (2025). Enhancing business intelligence through NLP and contextual AI synergy. Engineering Research Express, 7(2), 025279.

Rajan, A., Essa, N., Mohammed, N., Rashed, O., Siwe, A. T., Ravikumar, R., & Hassouna, M. (2025). Machine Learning for Supermarket Sales and Customer Segmentation. In 2025 10th International Conference on Information Technology Trends (ITT), pp. 130-135). IEEE.

Seidler-de Alwis, R. (2026). Qualitätssicherung im CI-Research mit KI-Tools. Information–Wissenschaft & Praxis.

Xu, J., Zhang, B., Li, H., Lu, J., & Yu, Y. (2026). From optimization to disruption: The role of AI-driven innovation in shaping firms' market competitiveness. Journal of Engineering and Technology Management, 80, 101962.

RELATIONSHIP BETWEEN DATA GOVERNANCE AND ARTIFICIAL INTELLIGENCE GOVERNANCE

Martin Zelenka

Prague University of Economics and Business
Faculty of Informatics and Statistics
martin.zelenka@vse.cz

DOI: 10.35011/IDIMT-2026-175

Keywords

Artificial intelligence governance; data governance; information management; organizational design

Abstract

The increasing adoption of artificial intelligence in organizations has intensified the need for governance mechanisms ensuring accountability, transparency, and compliance, while data governance remains a foundational capability for managing data assets. However, the relationship between data governance and artificial intelligence governance is conceptualized inconsistently in recent literature. This study addresses this gap through an interpretive literature review using inductive qualitative coding, the analysis identifies two overarching approaches to this relationship: independent governance domains and dependent governance domains. These approaches are further structured into seven patterns. The results show that the relationship between the two domains forms a continuum defined by the type of influence and degree of integration. This study identifies two opposing groups of factors explaining this variation: separation is associated with differences in governance objects, regulatory fragmentation, risk categories, and organizational roles, while integration is driven by the technical dependency of artificial intelligence on data, the role of data in related risks, and the operational realization of governance through shared data processes across the lifecycle. These findings provide a structured basis for understanding and designing governance arrangements in organizations.

1. Introduction

The increasing adoption of artificial intelligence (AI) in organizational contexts has intensified the need for governance mechanisms ensuring reliability, accountability, and compliance (Floridi et al., 2018; Camilleri, 2024). As AI systems are inherently dependent on data, the relationship between data governance and AI governance has become a critical issue for both research and practice (Mohammed, 2024). However, there is still no common theoretical framework connecting these domains into a unified governance approach.

Data governance is commonly defined as the specification of processes, roles, and accountabilities ensuring appropriate behaviour in the creation, storage, and use of data (Khatri & Brown, 2010; Otto 2011). Recent research positions it as a continuous organizational capability supporting transparency, trust, and alignment with business objectives (Acev et al., 2025).

AI governance refers to the system of policies, processes, roles, and controls guiding the development, deployment, and use of AI systems (Papagiannidis et al., 2023). It focuses on accountability, risk management, and compliance, particularly in relation to ethical and societal impacts (Kaur et al., 2022). Recent contributions further emphasize the operationalization of governance across the AI lifecycle, including its integration into technical infrastructures (Palumbo et al., 2025). This perspective remains relevant even where organizations rely on externally developed AI models, as governance increasingly concerns deployment contexts and operational use (Schneider et al., 2024; Morley et al., 2023). AI governance frameworks remain insufficiently consolidated and often evolve separately from data governance, despite their strong interdependencies (Batool et al., 2023; Mohammed, 2024). As a result, the relationship between the two domains is conceptualized inconsistently, and their alignment in organizational practice remains unclear.

This paper addresses this gap by analysing how the relationship between data governance and AI governance is conceptualized in recent literature. The study aims to (i) identify patterns describing their relationship, (ii) analyse the reasons for their integration and separation, and (iii) derive implications for organizational governance.

2. Methodology

The study adopts an interpretive literature review with inductive qualitative coding, following established guidelines for transparency and concept-centric analysis (Webster & Watson, 2002; vom Brocke et al., 2009). Coding was inspired by Gioia et al. (2013) but simplified.

The literature search was conducted in the Web of Science database, while Google Scholar served only as an exploratory supplementary source for terminology validation and identification of emerging themes. The review focused on publications from 2023 onward, reflecting the rapid organizational adoption of large language models such as ChatGPT and the increasing attention to accountability, risk management, and governance structures (Schneider et al., 2024).

Search query combined key terms related to both domains (AI governance and data governance) and used Web of Science integrated functionality to search in Topic (search field that scans the title, abstract, and keywords): *"algorithmic governance" OR "AI governance" OR "artificial intelligence governance" OR "model governance" (Topic) AND "data governance" OR "data management" OR "data management governance" OR "data stewardship" (Topic).*

The literature search conducted in the Web of Science database resulted in a total of 51 records. The search was performed in March 2026. To refine the sample, a set of exclusion criteria was applied. Firstly, 14 papers were excluded as they did not address the topic of AI governance. Secondly, 1 paper was removed due to retraction. Finally, 12 papers were excluded because they did not establish a clear relationship to data governance. After applying these criteria, the remaining 24 publications formed the final dataset. Although limited in size, the dataset was considered sufficient for interpretive theory-building aimed at conceptual pattern identification rather than statistical generalization (Webster & Watson, 2002).

The final dataset was subjected to a qualitative analytical process aimed at identifying reoccurring conceptualisations of the relationship between data governance and AI governance. Each paper was reviewed to identify relationships between data governance and AI governance, their underlying reasoning, and implications for practice. The analytical process followed an inductive coding approach inspired by the abstraction logic proposed by Gioia et al. (2013). The first stage involved identifying passages describing explicit or implicit relationships between data governance and AI governance. These passages were coded using concepts closely reflecting the terminology used in the

original publications (e.g., governance prerequisite, shared governance processes, separate accountability structures). Coding was performed by a single researcher and iteratively refined through repeated comparison between emerging categories and original source material to improve conceptual consistency. In the second stage, similar concepts were iteratively compared and aggregated into seven recurring relationship patterns representing distinct forms of interaction between data governance and AI governance. These patterns constituted the primary analytical categories of the study. In the final stage, the identified patterns were further abstracted into two overarching governance perspectives.

3. Results

3.1. Identified Patterns in the Relationship between Data Governance and AI Governance

The analysis identified two overarching structural perspectives on the relationship between data governance and AI governance, namely independent and dependent governance domains. The *independent perspective* comprises four patterns: *(1) foundational influence, (2) causal influence, (3) enabling or constraint influence, and (4) parallel complementary domains*. These patterns treat data governance and AI governance as distinct domains with separate responsibilities, where relationships occur through varying forms of influence rather than integration. Details are presented in Table 1.

The *dependent perspective* includes three patterns reflecting increasing levels of integration: *(5) component integration, (6) operational integration, and (7) unified lifecycle governance*. Here, data governance becomes progressively embedded within AI governance, evolving from a supporting component to integrated operational processes and, ultimately, unified governance across the AI lifecycle.

The patterns reveal two key dimensions: the *type of influence* (P1–P4) and the *degree of integration* (P5–P7), forming a continuum from loosely coupled to fully unified governance approaches. Full support denotes sources explicitly conceptualizing the relationship in a manner corresponding to the identified pattern. Partial support indicates indirect, implicit, or domain-specific alignment.

Table 1. Summarizes the mapping of the identified patterns to the analysed sources

(N)/Definition/1st stage examples	**Full/partial support (*cited example*)**
1. Data governance acts as a precondition layer for AI governance. A: “DG must exist before AI systems function responsibly” B: “DG forms prerequisite layer for ethical AI implementation”	Full: (Aguilar et al., 2026; A:Ai et al., 2026; Biswas, 2025; Garouani et al., 2025; Huang et al., 2025; Muallidin, 2026; B*:*Nakano, 2026; Palumbo et al., 2025; Pasas-Farmer, 2025; Perdana, 2025; Schulz, 2026); Partial: (Atoum, 2025; Billones et al., 2025; Ingber, 2025; Nguyen, 2026; Thuraisingham, 2024; Wei, 2025)
2. Data governance is not only enabling but deterministic for *AI*. A: “Poor DG directly produces AI/algorithmic risk” B: „Data extraction practices causally reproduce inequitable AI outcomes”	Full: (Biswas, 2025; A:Perdana, 2025; B*:*Schulz, 2026); Partial: (Atoum, 2025)
3. Data governance determines what AI governance can achieve (effects environment). A: “DG improves AI-enabled systems without implying direct dependency” B: “Poor DG constrains AI performance”	Full: (Aguilar et al., 2026; Ai et al., 2026; Alrasbi & Ariaratnam, 2026; Billones et al., 2025; Francu & Sava, 2025; James, 2023; A:Khan, 2024; McCormack, 2025; Muallidin, 2026; Nguyen,2026; Pasas-Farmer, 2025; Perdana, 2025; Primec et al., 2026; Syniehubov, 2025; Thuraisingham, 2024; Wei, 2025); Partial: (Ingber, 2025)

(N)/Definition/1st stage examples	Full/partial support (*cited example*)
4. Data governance and AI governance are separate but coordinated domains. A:" Multiple governance regimes coexist around AI" B:" AI governance requirements coexist alongside digital governance reforms"	Full: (Francu & Sava, 2025; Khan, 2024; McCormack, 2025; Nguyen, 2026; A: Primec et al., 2026; B: Syniehubov, 2025)
5. Data governance is embedded as a sub-element within AI governance. A: "DG is operationalized as one dimension inside AI governance" B:" DG integrated into AI monitoring architecture"	Full: (Alrasbi & Ariaratnam, 2026; Nakano, 2026; A:Pasas-Farmer, 2025, B: Palumbo et al.)
6. AI governance is executed through data governance practices. A: "AI governance implemented through existing governance mechanisms" B:" Governance implemented through operational pipelines"	Full: (A: Primec et al., 2026, B:Aguilar et al., 2026; Nakano, 2026; Perdana, 2025); Partial: (Francu & Sava, 2025)
7. A single, end-to-end governance model. A:" Governance converges across data, code, metadata, and models" B: "Framework for embedding AI ethics across the entire data lifecycle"	Full: (A: Garouani et al. 2025; B: Huang et al., 2025); Partial: (McCormack, 2025; Muallidin, 2026)

Source: author

The wide variety of patterns suggests that there is not convergence to some common attitude. Therefore, it makes sense to dive into analysis of factors causing such divergence.

3.2. Factors shaping the separation and integration

The literature identifies a set of opposing factors that explain why data governance and AI governance are either treated as separate domains or increasingly integrated. These factors can be grouped into those *supporting separation* and those *driving integration*, reflecting different conceptual, regulatory, organizational, and technical perspectives (Table 2).

The identified factors suggest an asymmetry between drivers of separation and integration. Factors supporting separation are predominantly associated with regulatory requirements, conceptual distinctions, and organizational specialization, whereas factors driving integration emerge more frequently from technical dependencies and practical governance needs across AI lifecycles. This may indicate that governance convergence is increasingly shaped by operational realities, while separation often reflects existing institutional arrangements and evolving regulatory frameworks. Evidence from regulated contexts such as finance, healthcare, and public administration suggests stronger emphasis on accountability, compliance, and risk management, although this does not consistently translate into a single preferred governance configuration. Instead, governance arrangements appear to be influenced more strongly by AI maturity, technical interdependencies, and organizational context than by sector alone. More recent contributions increasingly emphasize operational and lifecycle governance approaches, suggesting a possible shift toward stronger integration between data governance and AI governance. As AI governance matures, the balance between these forces may shift toward greater integration across governance domains.

Table 2. Factors influencing the relationship between data governance and AI governance

Governance dim.	Factors supporting separation	Factors supporting integration
Conceptual	Different governance maturity (Muallidin, 2026; Alrasbi & Ariaratnam, 2026; Khan, 2024) Different governance objects (Primec et al., 2026; Francu & Sava, 2025)	Lifecycle perspective on governance (Garouani et al., 2025; Huang et al., 2026; Palumbo et al., 2025)
Regulatory	Regulatory fragmentation (Francu & Sava, 2025; Khan, 2024; Primec et al., 2026); Different risk categories (Primec et al., 2026; Francu & Sava, 2025)	Shared accountability for AI risks and ethics originating from data (Biswas, 2025; Schulz & Loewen-Colón, 2026; Perdana et al., 2025)
Organizational	Organizational specialization of roles (Alrasbi & Ariaratnam, 2026; McCormack et al., 2025)	Cross-functional governance requirements across AI lifecycle (Garouani et al., 2025; Huang et al., 2025)
Technical-operational		Technical dependency of AI on data (Muallidin, 2026; Aguilar et al., 2026; Huang et al., 2025; Palumbo et al., 2025) Operationalization through shared processes (Aguilar et al., 2026; Nakano, 2026; Huang et al., 2025; Palumbo et al., 2025); Integrated technical infrastructures (Palumbo et al., 2025; Huang et al., 2025)

Source: author

3.3. Organizational impact

From an organizational perspective, the findings suggest that governance design should be treated as a contextual decision rather than a universal model. Organizations with low AI maturity, fragmented regulatory requirements, specialized governance roles, or separate data and AI infrastructures may benefit from maintaining partially separated or coordinated governance structures, supporting accountability, compliance, and clear role allocation (Khan, 2024; Primec et al., 2026). In contrast, organizations characterized by advanced AI adoption, integrated data ecosystems, cross-functional governance structures, or AI systems embedded in core processes may require stronger coordination or unified governance models to enable lifecycle oversight, risk management, and operational efficiency (Palumbo et al., 2025; Garouani et al., 2025).

The identified patterns therefore should not be interpreted as alternative best practices, but as governance configurations suitable under different organizational conditions. Regardless of the selected model, governance mechanisms should remain aligned with common objectives such as reliability, transparency, accountability, and trustworthiness (Aguilar et al., 2026; Biswas, 2025).

4. Discussion

Our research identified patterns in which data governance and AI governance are closely integrated, including component integration, operational integration, and unified lifecycle governance. These patterns view both domains as a single interdependent system across the AI lifecycle. This perspective is supported by Moh'd Al Dajeh (2024), who links transparency and accountability in AI to robust data governance, and by Mohammed (2024), who conceptualizes both domains as a symbiotic

relationship. Similarly, Camilleri (2024) shows that AI governance frameworks systematically incorporate data-related controls as core elements of responsible AI. These findings support integrated governance approaches, consistent with the dependent patterns identified in this study.

Conversely, our research also identified patterns where data governance and AI governance remain independent but interacting domains, particularly causal influence, enabling or constraint influence, and parallel complementary domains. These patterns reflect AI-specific risks, system autonomy, and evolving technological characteristics. Floridi et al. (2018) argue that AI governance must address broader ethical and societal concerns beyond data-related issues, while Hatta (2026) highlights governance challenges in advanced AI systems extending beyond traditional data governance assumptions. Together, these findings support the view that AI governance remains a distinct governance domain despite strong dependencies on data governance.

5. Conclusion

This paper contributes to the discussion on the relationship between data governance and AI governance by identifying seven patterns grouped into independent and dependent governance domains. These patterns are structured along two dimensions: the type of influence between governance domains and the degree of their integration. The findings suggest that governance configurations should not be treated as universal best practices but selected according to contextual factors such as regulatory environment, organizational structure, AI maturity, and technological architecture. The study thus provides a structured basis for designing governance arrangements in practice. At the same time, the results indicate that the field has not yet reached conceptual consensus, and efforts to define common governance models remain incomplete. Future research should examine whether preferred governance configurations differ systematically across industries and organizational contexts.

References

Acev, D., Biyani, S., Rieder, F., Aldenhoff, T. T., Blazevic, M., Riehle, D. M., & Wimmer, M. A. (2025). Systematic analysis of data governance frameworks and their relevance to data trusts. Management Review Quarterly. https://doi.org/10.1007/s11301-025-00545-1

Aguilar, R. M., Alayón, S., Torres, J. M., & Martín, C. A. (2026). Data-centric AI governance for responsible organizational value: Evidence from a European public administration. AI & Society. https://doi.org/10.1007/s00146-026-02978-y

Ai, K., Zhang, H., Bu, L., Călin, A. C., & Saleem, S. F. (2026). New development: Big data and artificial intelligence for smart public decision-making and advancing the SDGs. Public Money & Management, 1–5. https://doi.org/10.1080/09540962.2026.2639028

Alrasbi, O., & Ariaratnam, S. T. (2026). The Organizational Transformation of Artificial Intelligence in Smart Cities: An Urban Artificial Intelligence Governance Maturity Model. Urban Science, 10(1), 63. https://doi.org/10.3390/urbansci10010063

Atoum, I. (2025). Revolutionizing AI governance: Addressing bias and ensuring accountability through the holistic AI governance framework. International Journal of Advanced Computer Science & Applications, 16(2).

Batool, A., Zowghi, D., & Bano, M. (2023). Responsible AI Governance: A Systematic Literature Review (arXiv:2401.10896). arXiv. https://doi.org/10.48550/arXiv.2401.10896

Billones, R. K. C., Lauresta, D. A. S., Dellosa, J. T., Bong, Y., Stergioulas, L. K., & Yunus, S. (2025). AI Ecosystem and Value Chain: A Multi-Layered Framework for Analyzing Supply, Value Creation, and Delivery Mechanisms. Technologies, 13(9), 421. https://doi.org/10.3390/technologies13090421

Biswas, I. (2025). Relational accountability in AI-driven pharmaceutical practices: An ethics approach to bias, inequity and structural harm. Journal of Medical Ethics. https://jme.bmj.com/content/early/2025/09/09/jme-2025-110913.abstract

Brocke, J. vom, Simons, A., Niehaves, B., Niehaves, B., Riemer, K., Plattfaut, R., & Cleven, A. (2009). Reconstructing the giant: On the importance of rigour in documenting the literature search process. https://aisel.aisnet.org/cgi/viewcontent.cgi?article=1145&context=ecis2009

Camilleri, M. A. (2024). Artificial intelligence governance: Ethical considerations and implications for social responsibility. Expert Systems, 41(7), e13406. https://doi.org/10.1111/exsy.13406

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. Minds and Machines, 28(4), 689–707. https://doi.org/10.1007/s11023-018-9482-5

Francu, C. C., & Sava, Ș. (2025). Global Perspectives on Digital and AI Legislation: A Comparative Study of Data Protection, AI Governance, and Healthcare Innovations with a Focus on Romania. Proceedings of the International Conference on Business Excellence, 19(1), 2469–2481. https://doi.org/10.2478/picbe-2025-0191

Garouani, M., Ravat, F., & Valles-Parlangeau, N. (2025). Model Lake: A New Alternative for Machine Learning Models Management and Governance. In M. Barhamgi, H. Wang, & X. Wang (Eds), Web Information Systems Engineering – WISE 2024 (Vol. 15439, pp. 133–144). Springer Nature Singapore. https://doi.org/10.1007/978-981-96-0573-6_10

Gioia, D. A., Corley, K. G., & Hamilton, A. L. (2013). Seeking Qualitative Rigor in Inductive Research: Notes on the Gioia Methodology. Organizational Research Methods, 16(1), 15–31. https://doi.org/10.1177/1094428112452151

Hassani, H., Huang, X., & MacFeely, S. (2025). Mapping the evolution of data governance scientific research. Data & Policy, 7, e51. https://doi.org/10.1017/dap.2025.10014

Hatta, M. (2026). Several Issues Regarding Data Governance in AGI. In M. Iklé, A. Kolonin, & M. Bennett (Eds.), Artificial General Intelligence (Vol. 16057, pp. 239–249). Springer Nature Switzerland. https://doi.org/10.1007/978-3-032-00686-8_22

Huang, X., Kou, T., & Zhou, Q. (2026). Embedding AI ethics in the data lifecycle: A framework for enterprise AI governance. Technology in Society, 86, 103261. https://doi.org/10.1016/j.techsoc.2026.103261

Ingber, A. S., & Andalibi, N. (2025). Regulating Emotion AI in the United States: Insights from Empirical Inquiry. Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency, 167–180. https://doi.org/10.1145/3715275.3732014

James, A., Hynes, D., Whelan, A., Dreher, T., & Humphry, J. (2023). From access and transparency to refusal: Three responses to algorithmic governance. Internet Policy Review, 12(2). https://doi.org/10.14763/2023.2.1691

Kaur, D., Uslu, S., Rittichier, K. J., & Durresi, A. (2022). Trustworthy Artificial Intelligence: A Review. ACM Comput. Surv., 55(2), 39:31-39:38. https://doi.org/10.1145/3491209

Khan, A. (2024). The intersection of artificial intelligence and international trade laws: Challenges and opportunities. IIUMLJ, 32, 103.

Khatri, V., & Brown, C. V. (2010). Designing data governance. Communications of the ACM, 53(1), 148–152. https://doi.org/10.1145/1629175.1629210

Ledendal, J. (2026). AI Data Governance: Overlaps Between the AI Act and the GDPR. Law, Innovation and Technology. https://www.diva-portal.org/smash/record.jsf?pid=diva2:1996843

McCormack, L., Bendechache, M., Lewis, D., & Huyskes, D. (2025). Trust and transparency in AI: Industry voices on data, ethics, and compliance. AI & Society. https://doi.org/10.1007/s00146-025-02654-7

McGregor, S., & Hostetler, J. (2023). Data-Centric Governance (arXiv:2302.07872). arXiv. https://doi.org/10.48550/arXiv.2302.07872

Moh'd Al Dajeh, B. (2024). Artificial Intelligence Governance. Journal of Ecohumanism, 3(4), 300–313.

Mohammed, K. M. (2024). AI and Data Governance: A Symbiotic Relationship for Responsible Innovation. https://www.academia.edu/download/124867478/V13I1202415.pdf

Morley, J., Kinsey, L., Elhalal, A., Garcia, F., Ziosi, M., & Floridi, L. (2023). Operationalising AI ethics: Barriers, enablers and next steps. AI & Society, 38(1), 411–423. https://doi.org/10.1007/s00146-021-01308-8

Muallidin, I. (2026). Toward an algorithmic coloniality framework: Analyzing artificial intelligence governance in postcolonial Indonesia. Transforming Government: People, Process and Policy, 1–26.

Nakano, M. (2026). Bridging the AI Ethics Gap: A Tripartite Framework for Accountability, Implementation, and Governance. International Journal of Software Innovation (IJSI), 14(1), 1–14. https://doi.org/10.4018/IJSI.401497

Nguyen, M. T. (2026). Artificial intelligence in educational management: A conceptual review of leadership and governance implications. Journal of Applied Research in Higher Education, 1–15.

Otto, B. (2011). A morphology of the organisation of data governance. https://aisel.aisnet.org/ecis2011/272/

Palumbo, G., Guimarães, M., Carneiro, D., Marreiros, G., & Alves, V. (2025). Observability-Driven AI Governance: A Framework for Compliance and Audit Readiness Under the EU AI Act. New Trends in Disruptive Technologies, Tech Ethics and Artificial Intelligence (Vol. 1465, pp. 402–413). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-99474-6_36

Papagiannidis, E., Enholm, I. M., Dremel, C., Mikalef, P., & Krogstie, J. (2023). Toward AI Governance: Identifying Best Practices and Potential Barriers and Outcomes. Information Systems Frontiers, 25(1), 123–141. https://doi.org/10.1007/s10796-022-10251-y

Pasas-Farmer, S., & Jain, R. (2025). From discovery to delivery: Governance of AI in the pharmaceutical industry. Green Analytical Chemistry, 13, 100268. https://doi.org/10.1016/j.greeac.2025.100268

Perdana, A., Arifin, S., & Quadrianto, N. (2025). Algorithmic trust and regulation: Governance, ethics, legal, and social implications blueprint for Indonesia's central banking. Technology in Society, 81, 102838.

Primec, A., Belak, J., & Čufar, M. (2026). Artificial Intelligence as an Emerging Risk Dimension in Corporate Sustainability Reporting: A Legal and Governance Perspective. Sustainability, 18(5), 2278. https://doi.org/10.3390/su18052278

Schneider, J., Kuss, P., Abraham, R., & Meske, C. (2024, February 5). Governance of Generative Artificial Intelligence for Companies. arXiv.Org. https://arxiv.org/abs/2403.08802v5

Schulz, M., & Loewen-Colón, J. (2026). Preventing AI extractivism: The case for braiding indigenous data justice with ABS for stronger AI data governance. AI & Society. https://doi.org/10.1007/s00146-026-02931-z

Syniehubov, O. B., & Chernenko, O. (2025). European approaches to Digital Justice in central and eastern Europe and the Baltic states, with perspectives for Ukraine. Public Administration, 118, 9.

Thuraisingham, B. (2024). Trustworthy Artificial Intelligence for Securing Transportation Systems. Proceedings of the 29th ACM Symposium on Access Control Models and Technologies, 5–6. https://doi.org/10.1145/3649158.3657041

Weber, K., Otto, B., & Österle, H. (2009). One Size Does Not Fit All—A Contingency Approach to Data Governance. Journal of Data and Information Quality, 1(1), 1–27. https://doi.org/10.1145/1515693.1515696

Webster, J., & Watson, R. T. (2002). Analyzing the past to prepare for the future: Writing a literature review. MIS Quarterly, 26(2), xiii–xxiii.

Wei, C., Zhao, J., & Sun, L. (2025). Achieving regulatory alignment for E2E autonomous driving in China: A framework for tort liability and data governance. Computer Law & Security Review, 59, 106192. https://doi.org/10.1016/j.clsr.2025.106192

SOCIAL MEDIA, CREATIVITY AND INNOVATION: SYNERGIES AND TENSIONS

FROM RANKING TO REMOVAL: AI-DRIVEN SELECTION, MODERATION AND GOVERNANCE IN SOCIAL MEDIA

Antonín Pavlíček, František Sudzina, Jana Syrovátková, Richard A. Novak

Prague University of Economics and Business
Faculty of Informatics and Statistics
antonin.pavlicek@vse.cz, frantisek.sudzina@vse.cz, jana.syrovatkova@vse.cz, richard.novak.appletree@gmail.com

DOI: 10.35011/IDIMT-2026-185

Keywords

Artificial Intelligence; social media; recommender systems; content moderation; platform governance; privacy; digital services act; algorithmic transparency.

Abstract

Social media platforms increasingly use artificial intelligence not only to rank and personalize content, but also to detect, label, downrank, and remove content at scale. These functions are often analysed separately, although in practice they form one governance infrastructure that determines what users see, what remains hidden, and what becomes permissible within platform environments. This conceptual paper synthesizes current trends in AI-driven selection and moderation, focusing on their technical logic, socio-economic effects, ethical risks, and regulatory implications. It argues that AI improves scalability, relevance, and safety, but also deepens problems of opacity, privacy, bias, autonomy, and accountability. The paper further examines the European Union regulatory response, especially the Digital Services Act and the General Data Protection Regulation, and shows how transparency, risk mitigation, non-profiling alternatives, and user redress are becoming central requirements for platform governance. By integrating recommendation and moderation under a single analytical frame, the article contributes to a more coherent understanding of how AI reshapes the informational and normative architecture of social media.

1. Introduction

Social media platforms have moved from relatively open communication spaces to highly mediated environments in which artificial intelligence organizes visibility, reach, and the boundaries of acceptable participation. Users rarely encounter posts, videos, advertisements, or accounts in a neutral order. Instead, recommender systems rank content according to inferred preferences and predicted engagement, while moderation systems detect, flag, label, downrank, or remove content that platforms define as harmful, illegal, or otherwise undesirable (Pappalardo et al., 2024; Leblanc & Roux, 2025; Feerst, 2022).

Although recommendation and moderation are frequently studied as separate technical domains, they increasingly operate as two parts of the same governance architecture. Selection determines what is amplified; moderation determines what is restricted, obscured, or excluded. Together, these systems define the practical conditions under which communication becomes visible and legitimate on social media (Yu et al., 2024; Barral Martínez, 2023). This paper therefore treats AI-driven selection and moderation as interrelated mechanisms of platform governance rather than as isolated tools.

The argument is conceptual and synthetic. It does not test a narrow empirical hypothesis, but draws on current literature to clarify how these systems function, what consequences they generate, and why they matter for users, creators, platforms, regulators, and democratic communication. Our position is critical but not dismissive. AI is indispensable for managing scale, personalization, and safety on large platforms; however, its expanding role also intensifies concerns about opacity, asymmetrical power, limited contestability, privacy-invasive profiling, and weak public oversight (Walker & Milne, 2024; Pierson et al., 2023; Banchio, 2024).

The relevance of this topic has increased because social media platforms are now core infrastructures of personal communication, political debate, cultural production, and economic exchange. Ranking and moderation are no longer peripheral operations. They structure attention, influence content strategies, shape public salience, and determine the conditions under which expression becomes visible or invisible (Pierson et al., 2023; Pappalardo et al., 2024). Recommendation may improve relevance and user experience, but it may also support concentration, homogenization, polarization, and reduced diversity of exposure. Moderation may increase responsiveness and safety, but it may also produce false positives, inconsistent enforcement, contextual blindness, and chilling effects on lawful expression (Ge et al., 2024; Yu et al., 2024).

These issues intersect with broader debates on privacy, accountability, and platform power. Recommendation systems depend on extensive profiling and behavioral inference, while moderation systems rely on classification models, policy taxonomies, enforcement pipelines, and limited transparency toward users (Walker & Milne, 2024; Feerst, 2022). The core question is therefore not only whether AI works efficiently, but whether it operates in ways that remain legitimate, proportionate, contestable, and compatible with user rights. The European Union provides a particularly important regulatory context because the Digital Services Act (DSA) and the General Data Protection Regulation (GDPR) place transparency, risk mitigation, data protection, and procedural safeguards at the center of platform governance (European Commission, n.d.; Debiasi et al., 2025).

The paper proceeds as follows. Section 2 discusses AI-driven selection and recommender systems. Section 3 analyses AI-assisted moderation as a socio-technical process extending from detection to enforcement. Section 4 integrates selection and moderation as joint visibility-governance mechanisms. Section 5 examines socio-economic and ethical implications. Section 6 focuses on privacy and data governance. Section 7 situates the discussion in the EU regulatory framework. The final sections discuss the broader implications and conclude that social media platforms should be understood as algorithmically governed environments in which AI increasingly shapes both visibility and legitimacy.

2. AI-Driven selection on social media

AI-driven selection is one of the defining operating principles of contemporary social media. Instead of presenting content chronologically, platforms use recommender systems to rank, filter, and personalize posts, videos, accounts, groups, advertisements, and search results according to predicted relevance and expected engagement (Pappalardo et al., 2024; Mascari et al., 2025). Users therefore

do not simply navigate social media; they encounter an environment continuously curated by algorithmic inference.

At the technical level, recommender systems combine user data, item characteristics, interaction histories, and contextual signals. Inputs may include explicit indicators such as follows, likes, subscriptions, or ratings, but also implicit behavioral traces such as viewing time, pauses, scrolling rhythm, click-through patterns, repeated exposure, and abandonment signals (Masciari et al., 2025). Recent systems also use deep learning, contextual modeling, and transformer-based architectures to process large and rapidly changing streams of content. The technical sophistication is considerable, but so is the capacity of these systems to generate social effects beyond simple matching between users and content.

For this reason, the analysis must move from recommendation accuracy to recommendation impact. Pappalardo et al. (2024) identify a broad set of recommender-system outcomes, including concentration, content degradation, discrimination, diversity loss, echo chambers, filter bubbles, homogenization, polarization, radicalization, and volume effects. This perspective is crucial because recommendation systems do not merely reflect preferences; they also shape, stabilize, and sometimes narrow them over time. Repeated exposure contributes to habit formation, salience perception, and behavioral steering. Algorithmic selection can therefore influence not only what users consume, but also what they consider relevant, normal, or socially important.

Selection is consequently a form of platform governance. Ranking algorithms do not usually prohibit speech, but they exercise power by deciding what becomes prominent and what remains marginal, obscure, or practically absent from attention (Pierson et al., 2023). Selection governs informational opportunity structures: it affects which creators gain reach, which topics attract engagement, which narratives become socially prominent, and which commercial messages are monetized. Recommender systems are thus infrastructural decision-makers embedded in everyday digital communication rather than neutral background tools.

The literature on trustworthy recommender systems adds important normative criteria. Ge et al. (2024) argue that recommendation quality cannot be assessed only through predictive performance, because systems that are accurate but opaque, unfair, privacy-invasive, fragile, or difficult to control may still undermine trust. Explainability, fairness, privacy, robustness, and controllability are therefore central dimensions of trustworthy recommendation. In social media contexts, where users are exposed to continuous behaviorally optimized streams, these dimensions are not optional technical refinements but conditions of legitimate platform design.

Explainability is especially important. When users do not understand why a post, advertisement, account, or topic appears in their feed, they may perceive the system as arbitrary or manipulative, particularly when recommendations are socially sensitive or unrelated to expressed interests. Lack of explanation also weakens autonomy because users cannot interpret, correct, or challenge the logic of curation. Controllability is similarly relevant. Users may wish to reduce personalization, modify ranking parameters, escape repetitive exposure, or select a non-profiling feed. Without such options, platforms retain unilateral authority over visibility.

Privacy is equally central. Recommendation depends on data extraction and behavioral inference. The gains of personalization are often inseparable from profiling practices that users only partially understand and rarely control (Walker & Milne, 2024; Ge et al., 2024). This creates a structural tension: the more platforms optimize relevance, the more they rely on detailed traces of user activity. Privacy is not an external constraint added to recommendation systems after design; it is part of their operational foundation. The governance of selection is therefore also a governance of data.

At the socio-economic level, AI-driven selection is tied to platform business models. Recommendation systems support engagement, retention, advertising value, and creator dependence. When optimization criteria privilege attention and monetizable interaction, platform interests may diverge from users' long-term interests and from democratic values. Systems optimized for engagement may privilege emotionally salient, repetitive, or polarizing content even when this weakens informational diversity or public deliberation (Pappalardo et al., 2024). This does not make recommendation inherently harmful, but it requires evaluation beyond technical performance alone.

3. AI-Assisted moderation: From detection to enforcement

AI-assisted moderation is the second major operational layer of platform governance. If recommendation systems determine what is amplified, moderation systems determine what is restricted, labeled, downranked, removed, or referred for further review. Moderation is therefore not merely a reactive safety mechanism. It is the process through which platforms define acceptable participation, manage risk, and enforce normative boundaries (Singhal et al., 2023; Yu et al., 2024). Its growing automation reflects the scale and speed of digital communication, where exclusively human adjudication would be practically impossible (Feerst, 2022).

Moderation should be understood as a pipeline rather than as a single decision. Singhal et al. (2023) distinguish community guidelines and terms of service, violation and abuse detection, and policy enforcement. AI can enter each stage with different degrees of autonomy. It may identify suspicious content for human review, prioritize queues, support classification, estimate risk, recommend sanctions, or directly trigger labels, removals, restrictions, or account actions (Yu et al., 2024). A precise analysis must therefore distinguish detection support, decision support, and partially or fully automated enforcement.

At the detection stage, AI systems process text, images, video, metadata, user reports, network behavior, and dissemination patterns. They may identify hate speech, offensive language, misinformation, manipulated media, harmful imagery, spam, harassment, or coordinated abuse. The literature describes a mixed ecology of heuristic methods, rule-based systems, machine-learning models, deep neural networks, and transformer-based language models (Singhal et al., 2023). Detection is therefore increasingly probabilistic, data-intensive, and context-dependent rather than purely rule-bound.

However, detection is only the beginning. AI outputs are often probability scores rather than definitive judgments. A post may be classified as likely hateful, misleading, or unsafe, but the platform must still assess the type of violation, severity, jurisdiction, user history, proportionality of response, and possibility of error. Moderation is consequently a socio-technical process, not a narrow classification task (Yu et al., 2024; Feerst, 2022). Metrics such as precision and recall are useful but insufficient because they cannot settle normative questions of fairness, legitimacy, or proportionality.

The enforcement stage makes this tension clear. Singhal et al. (2023) distinguish hard moderation from soft moderation. Hard moderation includes removal of content or suspension of accounts. Soft moderation includes warning labels, reduced distribution, quarantining, age-gating, demonetization, or temporary restrictions. This distinction matters because soft moderation may reshape visibility without being obvious to users. AI therefore does not simply decide what is allowed; it increasingly influences the mode and intensity of intervention.

Degrees of automation further complicate accountability. Yu et al. (2024) show that research often conflates different forms of AI-assisted moderation, although systems vary greatly in the discretion they exercise and in the human oversight they preserve. Some tools merely surface potentially

problematic material; others perform parts of adjudication more independently. This affects user perception, institutional responsibility, and the location of accountability in human-machine workflows. When decisions emerge from layered systems, responsibility may become diffuse.

Human judgment remains indispensable because moderation routinely involves context, ambiguity, multilinguality, irony, satire, political contestation, and culturally specific meaning (Barral Martínez, 2023). What counts as harassment, hate speech, extremism, incitement, or misinformation often depends on audience, setting, intention, and legal context. Automation can support triage and consistency, but it cannot remove the interpretive character of moderation without increasing the risk of overreach, under-enforcement, or discriminatory outcomes.

This limitation is institutional as well as technical. Feerst (2022) argues that many problems attributed to AI moderation are rooted in the underlying subjectivity of moderation itself. Societies often disagree about the normative boundaries of acceptable speech. AI can accelerate enforcement, but it cannot eliminate disagreement about what should be enforced. Meaningful appeals, human review, and procedural safeguards are therefore not secondary additions. They are legitimacy mechanisms that allow contextual correction, proportionality, and contestability (Yu et al., 2024; Barral Martínez, 2023).

4. Selection and moderation as joint governance mechanisms

Recommendation and moderation are usually associated with different platform functions. Recommendation is linked to personalization, engagement optimization, and discovery; moderation to safety, rule enforcement, and harmful-content control. Yet from a governance perspective this separation is increasingly misleading. Both functions classify content, prioritize signals, and intervene in visibility. One determines what is elevated; the other determines what is reduced, labeled, restricted, or removed. Together, they form the operational core of AI-mediated platform governance (Pierson et al., 2023; Barral Martínez, 2023).

The conceptual link is visibility management. Recommendation systems influence whether content becomes visible, prominent, monetizable, and socially salient. Moderation systems influence whether content remains visible, is framed by warnings, loses distribution, or disappears from the platform. In both cases, the central output is not merely a binary decision, but a calibrated intervention in visibility (Barral Martínez, 2023). Social media governance is therefore exercised not only through deletion, but also through ranking, suppression, friction, labeling, and limited discoverability.

This shifts analysis beyond a narrow focus on censorship. A platform can profoundly shape discourse without formally banning speech. By prioritizing some content, suppressing borderline material, recommending particular creators, or reducing discoverability through soft moderation, platforms structure the communicative opportunity space. What matters is not only whether content remains technically available, but whether it is likely to be encountered, amplified, monetized, trusted, or socially validated.

AI intensifies this process because selection and moderation operate through large-scale inference. Both systems depend on behavioral traces, predictive models, policy taxonomies, optimization goals, and proprietary infrastructures. Their operation is often opaque to users and difficult for external observers to audit (Pierson et al., 2023; Banchio, 2024). Platform power is therefore exercised not only through explicit rules, but also through technical architectures that determine how rules are operationalized and how attention is distributed.

This joint architecture creates several normative consequences. First, it increases informational asymmetry between platforms and users. Platforms know far more than users about why content is

promoted, restricted, or removed, which weakens autonomy and contestability. Second, it normalizes less visible interventions such as downranking, demonetization, or reduced discoverability, which may seriously affect creators, social movements, and public debate. Third, it complicates accountability because harmful outcomes may result from the accumulation of small algorithmic interventions rather than from a single identifiable decision.

European debates on platform regulation reflect these concerns. Barral Martínez (2023) notes that AI-based filtering tools are increasingly central to content moderation in the EU, but their use raises risks of over-blocking, lawful-content suppression, and inadequate safeguards for fundamental rights. These risks are not limited to direct removals. Legal uncertainty, reputational pressure, and risk-averse enforcement may push platforms toward precautionary suppression of grey-area content. In such cases, recommendation and moderation converge into a shared logic of visibility control.

The deeper problem is the tension between scalability and legitimacy. Platforms need automated systems because digital communication is too large and fast for purely human governance. Yet the more platforms rely on AI to manage visibility, the more difficult it becomes to ensure fairness, explanation, proportionality, and redress. Automation can optimize throughput, but legitimacy depends on whether users can understand and challenge decisions that affect their participation. Governance by algorithm is therefore also a question of institutional design.

5. Socio-Economic and Ethical implications

The joint use of AI-driven selection and moderation has consequences that extend far beyond technical efficiency. These systems allocate attention, produce commercial value, structure creator opportunities, define acceptable participation, and shape trust in platform environments. Their evaluation must therefore include user autonomy, platform concentration, informational diversity, privacy, bias, and accountability, not only model performance (Walker & Milne, 2024; Pierson et al., 2023).

A central socio-economic implication is the platformization of attention. Recommendation systems are embedded in business models that reward engagement, retention, and monetizable interaction. They help users discover content, but they also steer exposure toward platform-defined goals. Content that is emotionally charged, repetitive, or behaviorally effective may circulate more easily than content that supports deliberation or diversity (Pappalardo et al., 2024). The problem is not personalization itself, but the normative defensibility of the optimization criteria governing socially central infrastructures.

User autonomy is directly affected. In AI-mediated environments, individuals do not choose against a neutral background. Ranked feeds, predicted preferences, prompts, notifications, frictionless sharing, and default settings continuously guide attention and behavior. These systems may improve convenience, but they may also narrow encountered perspectives and make the conditions of choice opaque (Ge et al., 2024). Autonomy is weakened not only by coercion, but also by invisible curation that shapes preference formation without meaningful awareness or control.

Trust is also at stake. Platforms depend on user confidence that ranking, moderation, and data practices are minimally fair and intelligible. A system may be efficient from the platform perspective while appearing arbitrary, manipulative, or discriminatory to users, especially when changes in reach, recommendations, or enforcement are not explained (Ge et al., 2024; Walker & Milne, 2024). Trust is particularly vulnerable where users experience sudden loss of visibility, inconsistent enforcement, or moderation outcomes that ignore context and intent.

Bias and uneven impact are persistent ethical concerns. AI systems are trained on data, policy assumptions, and linguistic patterns that may reproduce existing inequalities. Research and legal analysis identify recurring risks of over-moderation, under-moderation, discriminatory enforcement, and context-insensitive classification (Banchio, 2024; Barral Martínez, 2023). Algorithmic harm is therefore not limited to technical error. At platform scale, it may become unequal participation, distorted representation, or the silencing of already vulnerable voices.

Opacity compounds these risks. Users often cannot know why some content is promoted, why other content disappears from circulation, or why a moderation outcome occurs. This weakens procedural fairness and insulates platform authority from scrutiny (Banchio, 2024; Pierson et al., 2023). Opaque systems are ethically problematic not only because they may be wrong, but because affected users may be unable to understand, challenge, or correct them.

Power concentration is another consequence. Large social media platforms increasingly function as gatekeepers of digital communication. When visibility, monetization, participation, and reputation depend on proprietary AI systems, users, creators, media organizations, and public institutions become dependent on infrastructures they do not control. This dependence affects market power, creator livelihoods, news distribution, political communication, and cultural visibility (Pierson et al., 2023). Platform governance is thus also a matter of economic coordination and institutional dominance.

The key tension is again scalability versus fairness. AI enables platforms to process massive volumes of content, personalize experience, and respond to risks rapidly. Yet scaling selection and moderation does not automatically scale legitimacy. Small model errors, biased data, or opaque rules can be applied to millions of users simultaneously (Yu et al., 2024). The gains of automation must therefore be weighed against the normative costs of reduced agency, limited oversight, and unequal distribution of harm.

6. Privacy and data governance

Privacy and data governance are constitutive features of AI-driven social media governance. Both recommendation and moderation depend on extensive data collection, inference, classification, and retention. Privacy should therefore not be treated as an external compliance constraint imposed on otherwise neutral technologies. It is embedded in the operational logic of platforms that personalize experience, optimize advertising, assess risk, and enforce rules through data-intensive systems (Walker & Milne, 2024).

Recommendation systems illustrate this interdependence. Personalized ranking transforms behavioral traces--clicks, viewing time, sharing patterns, engagement histories, and inferred preferences--into probabilistic judgments about future relevance and susceptibility to engagement. The more refined the personalization, the more extensive the underlying data infrastructure tends to become. This creates a tension between utility and data minimization: content relevance is often optimized through precisely the behavioral inference that raises privacy concerns (Walker & Milne, 2024; Ge et al., 2024).

Moderation systems raise similar issues, although they are less visible in public debate. Automated moderation may process not only the content of a single post, but also metadata, account history, user reports, prior enforcement records, network patterns, and signals of coordinated behavior. These systems classify content, but they may also classify users, relationships, and risk profiles. The privacy question is therefore not only what data are collected, but how data are interpreted and what

administrative consequences follow. Once user data enter moderation pipelines, privacy intersects with due process and procedural fairness.

The relationship between profiling and user rights is particularly significant. Recent guidance on the DSA-GDPR interplay argues that behavioral analysis used for prediction in recommender systems can constitute profiling under Article 4(4) GDPR, and that the presentation of specific content may in some cases have serious consequences for individuals (StreamLex, 2025). Ranking is therefore not merely a neutral delivery mechanism. When recommender outputs shape opportunities, economic exposure, or social outcomes, they enter a demanding legal and normative terrain.

Data governance includes decisions about what data are collected, how long they are retained, how they are combined, what inferences are drawn, and how automated outputs affect visibility and participation. Walker and Milne (2024) argue that responsible social media governance requires fairer treatment of consumer data and attention to both short-term and long-term harms. This is crucial where users do not understand the extent to which their data support invisible ranking, categorization, or risk scoring.

Transparency is therefore inseparable from privacy. If users cannot understand which categories of data are used, how data feed into recommendation or moderation, and what consequences follow from automated assessments, formal privacy rights remain weak in practice. The DSA-GDPR discussion emphasizes clear information, explanation of main recommender parameters, proportionality, and consistency between content governance and data-protection principles (Debiasi et al., 2025; StreamLex, 2025). Opacity in data governance undermines not only privacy, but also trust in platform authority.

Data minimization is difficult in environments built around prediction and optimization. Under GDPR logic, platforms should process only what is necessary for specified purposes, yet AI systems often expand the range of relevant signals and extract secondary inferences from behavior. What is technically useful may exceed what is normatively justified. This is especially important for moderation and notice-handling systems, where additional identification or disclosure should remain necessary and proportionate (StreamLex, 2025).

Meaningful choice is equally important. The DSA framework places growing emphasis on non-profiling recommendation alternatives and on users' ability to understand and modify ranking parameters (European Commission, n.d.; Debiasi et al., 2025). If users are locked into profiling-based systems by design, personalization becomes less a service preference than a condition of participation. Responsible data governance must therefore combine internal controls with user-facing agency.

7. EU regulation and platform governance

The European Union has become a key regulatory context for AI-mediated social media governance. The DSA strengthens obligations concerning transparency, accountability, risk mitigation, advertising practices, recommender systems, content moderation, and user protection, with particularly demanding duties for very large online platforms and search engines (European Commission, n.d.). The regulation is directly relevant because it addresses the areas in which AI increasingly shapes platform environments.

One important contribution of the DSA is its treatment of moderation as a matter of due process rather than internal platform discretion alone. Platforms must provide clear and specific statements of reasons when content is removed or accounts are restricted, and users must have access to complaint-handling and dispute-settlement mechanisms (European Commission, n.d.). For AI-assisted moderation, the implication is that automation cannot remain entirely invisible. Where algorithmic

systems support enforcement outcomes, those outcomes increasingly need to be explainable, contestable, and procedurally grounded.

The DSA is also significant for recommender systems. It requires transparency about the main parameters used to rank and present content and, for very large platforms, stronger user control over personalization, including options that are not based on profiling (European Commission, n.d.). This treats ranking not simply as a product feature but as a governance function affecting the informational conditions of online participation. The architecture of the feed becomes an object of public concern.

The relationship between the DSA and the GDPR is central. Debiasi et al. (2025) explain that both regimes are built around transparency and fundamental rights, but the DSA extends these ideas from personal data processing to the broader ecosystem of content governance, recommender systems, advertising transparency, and platform design. The GDPR gives individuals rights to information about personal-data processing and certain automated processes; the DSA adds obligations concerning how platform systems affect what users see and how they are treated.

This complementarity produces a more operational concept of transparency. It is not enough to disclose that data are processed or that algorithms are used. Information must be intelligible, accessible, and actionable. Users must be able to understand the main logic of ranking, the reasons for moderation outcomes, the existence of personalization choices, and the possibilities for redress (Debiasi et al., 2025). In AI governance, transparency is therefore linked to agency rather than to formal disclosure alone.

The DSA also introduces stronger public oversight. The European Commission and national Digital Services Coordinators supervise compliance, while transparency reporting, statements-of-reasons databases, and systemic-risk obligations create new forms of scrutiny (European Commission, n.d.). This is important because private ordering by platforms is insufficient when digital infrastructures have large-scale societal effects. Recommendation and moderation are no longer governed only inside the platform; they increasingly involve public institutions, external review, and coordinated enforcement.

Very large platforms are especially important because their recommender systems and moderation infrastructures can affect democratic discourse, minors, advertising integrity, consumer protection, and the broader informational environment. The DSA therefore connects platform design with systemic-risk thinking. It focuses not only on isolated moderation errors or individual transparency failures, but also on aggregate effects on public communication (European Commission, n.d.; Barral Martínez, 2023). This aligns with the central argument of this paper: recommendation and moderation should be analysed as interlocking governance mechanisms whose cumulative impact extends beyond individual interactions.

Regulation does not remove all tensions. Stronger pressure to detect and mitigate risks may encourage precautionary or overbroad automation, creating conflicts between safety, freedom of expression, privacy, and procedural fairness. The DSA seeks to manage these conflicts through transparency, redress, and accountability, but effectiveness will depend on implementation, enforcement, audit capacity, and platform willingness to operationalize obligations proportionately (Debiasi et al., 2025; Banchio, 2024). The EU approach is nevertheless significant because it moves beyond voluntary self-regulation and treats platform AI as part of a digital constitutional order grounded in accountability and fundamental-rights protection.

8. Discussion

The preceding analysis shows that AI-driven recommendation and moderation are mutually reinforcing components of a broader governance architecture. Recommendation systems distribute attention and informational exposure; moderation systems define the boundaries of acceptable participation and visibility. Together, they determine not only what users are likely to encounter, but also what kinds of communication become amplified, marginalized, restricted, or removed (Pappalardo et al., 2024; Yu et al., 2024). AI in social media should therefore be understood less as a collection of isolated tools and more as an infrastructural mechanism of power.

A major implication is that technical performance is no longer the central issue. The crucial question is whether AI-mediated governance can be legitimate under conditions of large-scale automation, commercial optimization, proprietary control, and asymmetrical access to data. Existing research shows effects at individual, group, organizational, and ecosystem levels, including concentration, diversity loss, unequal enforcement, and shifts in collective communication patterns (Pappalardo et al., 2024; Yu et al., 2024). Public debate often focuses on isolated moderation failures, but this may obscure the broader structural issue: AI systems increasingly organize the conditions of public communication.

The tension between scalability and legitimacy remains unresolved. Platforms depend on AI because fully human governance is impractical at scale. Recommendation enables continuous personalization; moderation enables rapid filtering and enforcement. Yet the same features that make these systems operationally effective create legitimacy deficits. Opaque ranking, probabilistic enforcement, limited contestability, and concentrated decision-making power weaken agency and external accountability (Banchio, 2024; Pierson et al., 2023). Scale may be technologically necessary, but it does not provide normative justification.

The problem is not the use of AI as such. Nor are all forms of automation inherently suspect. The problem is that automation is often embedded in business models and governance arrangements that privilege engagement, risk management, and operational efficiency over transparency, contestability, and public accountability (Walker & Milne, 2024). Under such conditions, technically sophisticated systems may function well for platforms while failing broader standards of fairness, legitimacy, and democratic acceptability.

The literature also remains fragmented. Yu et al. (2024) note that moderation research often focuses on individual-level harms and removal of harmful content, while group-level outcomes and constructive forms of moderation are less developed. Similarly, recommender-system research has begun to map ecosystem impacts more systematically, but concepts, methods, and evaluative criteria remain dispersed (Pappalardo et al., 2024). Future research should connect individual, organizational, and systemic levels, especially in relation to platform power, collective rights, public discourse, and long-term changes in digital public space.

Regulatory developments point in a promising direction, particularly in Europe. The DSA and GDPR-related transparency framework recognize that recommendation, moderation, profiling, advertising, and user redress cannot be governed as isolated issues (Debiasi et al., 2025; European Commission, n.d.). However, legal duties will be effective only if they are supported by institutional oversight, technical auditability, meaningful user agency, and transparent implementation. Governance requires not only rules, but also capacities to inspect, contest, and correct platform systems.

A further implication concerns research design. Future studies should not evaluate recommender systems only through engagement or satisfaction, nor moderation systems only through removal accuracy. More useful designs would examine the whole visibility pipeline: how content is selected,

how borderline cases are classified, how interventions change distribution, how users understand these decisions, and whether appeals actually correct errors. Mixed methods are especially valuable because quantitative audits can identify structural patterns, while qualitative work can reveal how affected users, creators, moderators, and platform professionals interpret them. Comparative research across platforms and jurisdictions would also help distinguish general problems of AI governance from effects produced by specific business models or legal regimes. Such work is necessary if platform governance is to become empirically accountable rather than merely rhetorically transparent.

9. Conclusion

This paper has argued that AI-driven recommendation and moderation are the twin governance infrastructures of contemporary social media platforms. Recommendation systems organize visibility and exposure, while moderation systems organize restriction and enforcement. Together, they shape the informational and normative architecture of online participation (Pappalardo et al., 2024; Yu et al., 2024).

AI improves scalability, relevance, responsiveness, and platform safety, but it also intensifies concerns about opacity, privacy, profiling, bias, user autonomy, concentrated power, and accountability (Walker & Milne, 2024; Banchio, 2024). These challenges cannot be addressed through technical optimization alone. They require governance arrangements based on transparency, proportionality, contestability, data protection, independent oversight, and meaningful user control.

The EU regulatory framework, especially the DSA and GDPR, offers an important institutional response by treating recommender systems, moderation procedures, transparency duties, and user rights as interconnected components of platform governance (European Commission, n.d.; Debiasi et al., 2025). Its success will depend on implementation, enforcement, and the ability to make complex AI systems auditable and contestable in practice.

Social media platforms should therefore be analysed not merely as communication technologies, but as algorithmically governed environments in which AI increasingly determines both what becomes visible and what becomes legitimate. The central challenge for future research, regulation, and platform design is to ensure that this visibility governance remains efficient without becoming opaque, scalable without becoming unaccountable, and personalized without undermining autonomy and fundamental rights.

AI disclaimer

The authors used generative AI tools only as auxiliary writing support (e.g. for translation, language polishing and paraphrasing). All conceptual framing, argumentation, structure of the paper, and selection and interpretation of the cited literature were carried out by the authors.

References

Barral Martínez, M. (2023). Platform regulation, content moderation, and AI-based filtering tools: Some reflections from the European Union. JIPITEC – Journal of Intellectual Property, Information Technology and E-Commerce Law, 14(1), 211–211. https://nbn-resolving.de/urn:nbn:de:0009-29-57165

Debiasi, L., Florina P. (2025). The Right to Information Evolved: GDPR and the DSA's Transparency Landscape. Eipa. https://www.eipa.eu/blog/dsa-and-gdpr-transparency-obligations/

Feerst, A (2022). The use of AI in online content moderation. American Enterprise Institute. https://platforms.aei.org/wp-content/uploads/2022/09/The-Use-of-AI-in-Online-Content-Moderation.pdf

European Commission. (n.d.). The impact of the Digital Services Act on digital platforms | Shaping Europe's digital future. https://digital-strategy.ec.europa.eu/en/policies/dsa-impact-platforms

Ge, Y., Liu, S., Fu, Z., Tan, J., Li, Z., Xu, S., Li, Y., Xian, Y., & Zhang, Y. (2024). *A* Survey on Trustworthy Recommender Systems Journal of Artificial Intelligence Research, 71, 1137–1181 (arXiv:2207.12515). arXiv. https://doi.org/10.48550/arXiv.2207.12515

Leblanc, C., & Roux A. (2025). AI and Social Media: An Integrative Literature Review. Communication Management. ⟨hal-05041614⟩

Pappalardo, L., Citraro, S., Cornacchia, G., Nanni, M., Pansanella, V., Rossetti, G., Gezici, G., Giannotti, F., Lalli, M., Mauro, G., Barlacchi, G., Gambetta, D., Morini, V., Pedreschi, D., & Ferragina, E. (2024). A survey on the impacts of recommender systems on users, items, and human-AI ecosystems (arXiv:2407.01630). arXiv. https://doi.org/10.48550/arXiv.2407.01630

Pierson, J., Kerr, A., Cory, R. (Stephen, Fanni (Rosanna), Eveline), S. (Valerie, Milan (Stefania), & Zampedri (Giulia). (2023). Governing artificial intelligence in the media and communications sector. Alexander von Humboldt Institute for Internet and Society gGmbH. https://doi.org/10.14763/2023.1.1683

Banchio, P. R. (2024). Legal, Ethical and Practical Challenges of AI-Driven Content Moderation (SSRN Scholarly Paper 4984756). Social Science Research Network. https://doi.org/10.2139/ssrn.4984756

Masciari, E., Umair, A., & Ullah, M. H. (202). A Systematic Literature Review on AI-Based Recommendation Systems and Their Ethical Considerations. IEEE Access, 12, 121223–121241. https://doi.org/10.1109/access.2024.3451054

Singhal, M., Ling, C., Paudel, P., Thota, P., Kumarswamy, N., Stringhini, G., & Nilizadeh, S. (2023). SoK: Content Moderation in Social Media, from Guidelines to Enforcement, and Research to Practice. IEEE 8th European Symposium on Security and Privacy (EuroS&P), 868–895. https://doi.org/10.1109/EuroSP57164.2023.00056

StreamLex. (2025). DSA and GDPR interplay explained: A guide for online services. https://streamlex.eu/news/dsa-gdpr-compliance-guidelines/

Walker, K. L., & Milne, G. R. (2024). AI-driven technology and privacy: The value of social media responsibility. Journal of Research in Interactive Marketing, 35(1). Emerald Publishing. https://doi.org/10.1108/JRIM-02-2024-0072

Yu, Z., Otto, L., Assenmacher, D., & Wagner, C. (2024). A Systematic Review of the Effects of AI-Assisted Moderation on Individuals and Groups. Human-Machine Communication, 9(1), 167-188. https://doi.org/10.30658/hmc.9.10

USER BEHAVIORAL FACTORS INFLUENCING INFORMATION SECURITY ON SOCIAL NETWORK SITES

Michael Anderle, Martin Lukáš

Czech University of Life Sciences Prague
Faculty of Economics and Management
anderle@pef.czu.cz, lukas@pef.czu.cz

DOI: 10.35011/IDIMT-2026-197

Keywords

Information security; behavioral intentions; social network sites; information processing; heuristic processing; phishing.

Abstract

Social networking sites represent a key component of contemporary digital information ecosystems, where information security depends not only on technology but also on user behavior. This study examines how selected behavioral intentions influence the likelihood of experiencing information security incidents on social network sites. Using survey data from 151 active users and Pearson's chi-square analysis, the results show significant associations between security incidents and habitual behaviors such as password updating, personal information sharing, and active privacy management. In contrast, no significant relationship was found between formal security awareness training and attack experience. The findings highlight the importance of habitual, behavior-oriented approaches to information security management in digital environments.

1. Introduction

The social media revolution has completely altered how we trade, use, and communicate information. Even though we have just had them for the past 20 years, 61.4% of total global population uses them daily and they spend almost 15% of their walking lives on the social media platform, according to 2023 Global social media statistics on the data portal. Social media is an internet-based platform that facilitates users in connecting, creating, and sharing information in various formats, such as blogs, social networking sites (SNS), podcasts, and more. In today's data-driven world, the main challenge for social media platforms is ensuring information security. There were countless data breaches, including the 2019 Facebook event, where more than 530 million users' personal data became public, and the 2021 LinkedIn compromise affecting more than 90% of its user base (Purdue Global, 2023). On SNS, the concern of privacy has been rising since 2007 when platforms like Facebook changed their privacy settings to gain more control over user data. However, many experts believe that the most crucial aspect of the security system is not technology but user behavior. Numerous studies demonstrate that users are the biggest weakness of information system security. According to (Ali et al., 2020), 43% of data breaches are due to employees' behaviors. As to the Verizon 2021 record, 83% of the cybersecurity issues were caused by human factors. This study focuses specifically on

user-level information security, i.e., behaviors that influence the likelihood of individual users being exposed to security incidents while using social networking sites (SNS), rather than the technical security of the platforms themselves. SNS are understood as major interactive platforms such as Facebook, Instagram, LinkedIn, and similar services that enable user-generated content and interpersonal communication. While the findings are derived from SNS environments, some behavioral mechanisms may also apply to other digital platforms; however, this generalization is outside the primary scope of this study. The study does not differentiate between specific SNS platforms, as the focus is on general behavioral patterns across commonly used services.

2. Literature review

2.1. Information security principles and management

Information security (InfoSec) describes the techniques, instruments, and policies employed to protect private information from threats and unauthorized actions, ensuring confidentiality, integrity, and availability of the "CIA Triad". Confidentiality prevents unauthorized access through measures like encryption and Access Control Lists (ACLs), integrity safeguards data against erroneous modification, and availability ensures systems are accessible when needed (Integrity safeguards data against unauthorized modification, commonly using mechanisms such as cryptographic hashes (e.g., HMACs), checksums, and digital signatures). Effective InfoSec management operates through "the six Ps": planning, policy, programs, protection, people, and project management (Whitman & Mattord, 2019). Policy framework is essential, covering Enterprise Information Security Policy (EISP), Issue-Specific Security Policies (ISSPs), and System-Specific Policies (SysSPs). Furthermore, security technology serves as a critical layer. This includes firewalls, antivirus software, and more recently, Blockchain technology for distributed ledgers and Machine Learning for detecting suspicious user behavior in encrypted traffic. Cloud Security Posture Management (CSPM) and endpoint security have also progressed to defend against sophisticated zero-day threats.

2.2. Social network sites and algorithmic influence

Social networks are platforms enabling individuals and businesses to connect and exchange information. According to (DataReportal, 2022), there are over 4.74 billion SNS users worldwide. These platforms can be categorized into social connections (Facebook, Instagram), professional connections (LinkedIn), news/information (Reddit), communication, and multimedia sharing (YouTube, TikTok). The system is governed by an algorithmic set of instructions that filter and recommend content based on location, interests, and demographics. On platforms like Facebook, timing and previous interactions are prioritized to keep users engaged. However, this "need for fast" content delivery has raised concerns about internet addiction and user dependency on these systems.

2.3. Human behavior and psychological principles

User behavior describes how individuals interact with a website, service, or platform. Tracking and analyzing these behaviors helps evaluate user intentions and interests, which can be used to improve user experience or company strategy. Cognitive processes, feelings, and motives influence how users engage with information. Several psychological principles are critical for understanding activities:

- Cognitive Load: Limited user cognition requires simple interfaces to reduce strain.
- Visual Perception: Guides attention to key elements.
- Hick’s Law: More options increase decision time and risk paralysis.

- Gestalt Principles: Understanding proximity, resemblance, and closure is required to design coherent and ordered interfaces.
- Fitts's Law: The time to move to a target is influenced by its size and distance.
- Social Proof: Users are influenced by the actions and decisions of others, which can be used to promote desirable actions.
- Reciprocity: Users often react favorably to gestures of generosity, suggesting value should be provided up front.
- Anchoring: Initial knowledge or values influence future decisions and user impressions.

Human error is estimated to cause more than 95% of successful cyberattacks. In the digital era, time is an essential factor; more than 25% of phishing emails are designed to induce a sense of urgency to lure recipients into non-secure responses. According to Cowan's theory, people only attempt to explain unclear differences during problem recognition, and under pressure, they may fail to classify a perceived disagreement as an issue. Furthermore, rushing leads to increased risk-taking, and research by (Lin & Jia, 2023) claims that under limited time, people tend to make risky decisions. The Optimism Bias further complicates security; (West, 2008) and (Wiederhold, 2014) indicate that individuals tend to be optimistic and downplay the possibility of negative events, believing they are not in danger. These biases lead users to minimize the significance of cyberattacks affecting them.

2.4. Challenges, awareness and security policy

The primary challenge on SNS is balancing security with accessibility. While tools like Authentication (biometrics, passwords), Access Control (ACL, RBAC), and Encryption are used, human error remains the "weakest link". IBM in 2014 estimated that 95% of successful attacks are caused by human error, while Ali et al. (2020) note that 43% of data breaches result from employee behavior. (Herath et al., 2022) identify that awareness plays a vital role in preventing threats, yet (Ali et al., 2021) point out that high security requirements can act as an external stressor. Employees who experience "security stress" are more likely to inadvertently break policies. Individual differences such as procrastination, impulsiveness, and future thinking (Egelman & Peer, 2015; Hadlington, 2017) also correlate with the intention to adhere to security regulations.

2.5. Mobile device security behavior

The "Go Mobile" movement provides immense flexibility, but sensitive information stored on mobile devices is increasingly compromised by security risks affecting all mobile services and applications. According to (Wang et al., 2020), the authentication process is crucial for preventing unauthorized access to these delicate data. Existing methods are classified into four fundamental categories: knowledge-based (passwords and lock patterns), physiological biometrics (fingerprints, face, and iris), behavioral biometrics (gait, swiping, and hand gestures), and multi-factor authentication. Knowledge-based methods remain the most common but are susceptible to attackers who can steal secrets by observing user input or following fingerprints left on touchscreens. It is known that attackers can fake physiological biometric data to trick the system, and since these traits cannot be altered, such breaches are irreversibly insecure. Behavioral biometrics, such as signature patterns, are becoming more acceptable as they are non-intrusively gathered, yet they remain vulnerable to emulation by threat actors. While two-factor or multi-factor authentication increases security by integrating multiple metrics, it may decrease user convenience by requiring numerous inputs. Finally, research by Nisreen Ameen highlights significant gender and cultural disparities in workers' security behavior intentions, specifically regarding national cultural values and their impact on BYOD (Bring Your Own Device) security practices. Authentication factors can also include possession-based

methods (e.g., tokens, mobile devices, or passkeys), which are increasingly used in modern authentication systems, although they introduce risks related to device theft or compromise.

3. Methodology

The primary objective of this research is to analyze users' behavioral intentions and their impact on Information Security (InfoSec) within Social Network Sites. The study follows a quantitative research design utilizing primary data collection through a structured survey and statistical evaluation.

3.1. Research design and survey instrument

The research adopts a cross-sectional design using a structured questionnaire grounded in behavioral psychology and information security frameworks discussed in the literature review. To ensure data quality, the questionnaire was pilot-tested with five participants to identify and resolve ambiguities in wording. The finalized survey consisted of 12 questions covering demographics and behavioral scenarios on social networking platforms. Data collection was conducted between October and December 2024 over a period of 12 weeks. Respondents were recruited using purposive and snowball sampling via online distribution channels, including social networking platforms (e.g., Facebook, LinkedIn, Instagram) and university mailing lists. This approach enabled access to active SNS users but does not represent a fully random sample. The sample shows a higher proportion of university-educated respondents (53%), which reflects the primary distribution channels and should be considered when interpreting the results. The questionnaire included closed-ended items using ordinal frequency scales (e.g., “Never”, “Rarely”, “Usually”, “Everyday”) to measure behaviors such as password updates, information sharing, and privacy settings usage. The dependent variable (“experience of a security attack”) was measured as a binary self-reported outcome (Yes/No).

3.2. Sample selection and demographic analysis

The target population included active SNS users regardless of geographic location, though the primary reach was within Europe. A total of 151 valid responses were gathered. Non-probability purposive sampling was combined with snowball sampling to maximize the diversity of respondents. As shown in Table 1, the sample is balanced by gender, allowing a comparative analysis of behavioral intentions without significant gender bias. The heavy concentration of respondents in the 18–60 age group (91.4%) aligns with global statistics on the most economically and socially active SNS users.

3.3. Research questions

To systematically evaluate the impact of user behavior on security outcomes, the following research questions were established. These questions are designed to contrast specific behavioral intentions (independent variables) against the actual experience of a security attack (dependent variable):

- *RQ#1: Does the frequency of updating passwords have a significant relationship with the experience of being a victim of an information security attack?* This question focuses on technical maintenance and the "common sense" security advice of regularly rotating credentials. It aims to verify if traditional proactive measures still provide a statistically significant defense in the modern SNS landscape where automated attacks are prevalent.
- *RQ#2: Is there a significant relationship between the frequency of sharing personal information online and the likelihood of experiencing a security attack?* This focuses on information disclosure and user self-expression. Given that social networks encourage the

sharing of personal data, this RQ explores whether high levels of openness (e.g., posting statuses or personal photos daily) attract more malicious actors and phishing attempts.

- *RQ#3: Does the practice of disabling location services or restricting post visibility impact the experience of security threats?* This question addresses proactive privacy management. It examines whether users who actively use platform-specific privacy tools, such as restricting viewers or disabling GPS location, are less likely to be targeted by location-based or socially engineered attacks.
- *RQ#4: Does prior awareness of online threats or completion of security training significantly correlate with a reduction in experienced attacks?* This focuses on the role of institutionalized knowledge and Security Education, Training, and Awareness (SETA) programs. It aims to determine if formal security education successfully translates into safer outcomes or if "security stress" and behavioral biases override theoretical knowledge. The study does not differentiate between one-time and repeated security training, which may influence its effectiveness.

Table 1. Demographic profile of respondents (N=151)

Variable	Category	Frequency	Percentage
Gender	Female	66	43.7 %
	Male	70	46.4 %
	Prefer not to say	15	9.9 %
Age Group	Under 18	10	6.6 %
	18 - 26	60	39.7 %
	27 - 60	78	51.7 %
	60+	3	2.0 %
Education	University	80	53.0 %
	High school	57	37,7 %
	Middle school	14	9.3 %

Source: authors

These research questions test whether individual behavioral habits correlate with a lower risk of victimization in social networking environments.

3.4. Data processing and statistical apparatus

Quantitative data was exported from Google Forms to MS Excel for initial cleaning and then imported into IBM SPSS Statistics v.27. Categorical Data Analysis (CDA) was chosen as the primary framework. Given that the variables (e.g., "Frequency of updates" vs. "Experience of attack") are nominal or ordinal, the Pearson Chi-square test (X^2) is the most appropriate method to determine whether a significant association exists between them. For the analysis, the significance level (p-value) was set at $\alpha = 0.05$. This rigorous statistical approach ensures that any identified relationships are not due to random chance, providing a solid foundation for the subsequent results.

4. Results

The practical part of this study evaluates how users' behavioral intentions affect information security on SNS using a sample of 151 respondents. Data were evaluated using descriptive statistics and the Pearson Chi-square (X^2) test via SPSS.

4.1. Evaluation of user activities and trust

The survey examined the intensity of social media engagement over the last three months. As shown in Figure 1, the activity level is high; 28.7% of respondents use social platforms every day, and 50% use them usually (3-5 days per week), meaning more than 78% of participants are frequent users. This activity is directly linked to the volume of data shared online. Regarding behavioral intentions, 22.5% of surveyed users post personal images and statuses daily, while 51.7% post usually. In total, 74.2% of respondents frequently share personal information online. Interestingly, high engagement does not correlate with high trust. Only 21.9% of respondents entirely trust their social platform's security measures. As illustrated in Figure 2, a significant 51.7% of those surveyed maintain a neutral attitude or do not trust the platforms at all.

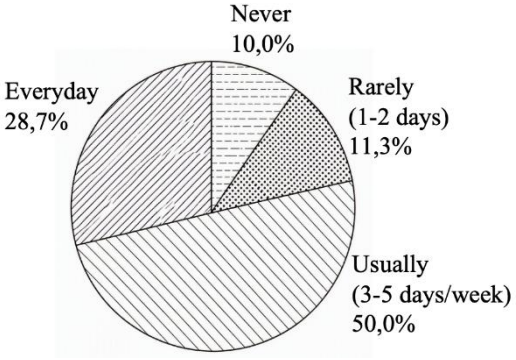


Figure 1. Frequency of using social networks (N=151)
Source: authors

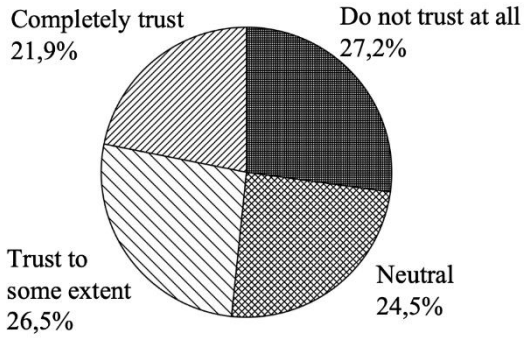


Figure 2. User's trust on their social platforms
Source: authors

Table 2. Personal information sharing and attack experience (N=151)

Frequency of sharing	Attack: No	%	Attack: Yes	%	Total	%
Everyday	15	44,2	19	55,8	34	100
Never	21	100	0	0	21	100
Rarely (1-2 days)	12	66,7	6	33,3	18	100
Usually (3-5 days/week)	40	51,3	38	48,7	78	100
Total	88	58,3	63	41,7	151	100

Source: authors

4.2. Statistical analysis of research questions (RQ#1–RQ#4)

The dependency between user behavioral intentions and information security attack experiences was tested using the Pearson Chi-square test (X^2). The significance level was set at $\alpha = 0.05$.

- RQ#1 (Password updates): The analysis showed a significant dependency between the frequency of updating passwords and being attacked ($p = 0.001,\ X^2 = 16.819, df = 3$). The data indicates that users who update passwords "Weekly" or "Monthly" report different attack frequencies compared to those who "Never" or "Rarely" do so.
- RQ#2 (Information sharing): A strong dependency was found regarding the frequency of sharing personal information online ($p < 0.001,\ X^2 = 19.929, df = 3$). As shown in Table 2, a critical finding is that respondents who "Never" share personal information reported zero security attacks.
- RQ#3 (Privacy settings): The practice of disabling location services or restricting post visibility also proved to have a significant relationship with attack frequency ($p = 0.004$,

$X^2 = 8.263, df = 1$). Users who actively used these privacy features reported different experiences.

- RQ#4 (Security awareness): Contrary to theoretical expectations, no significant relationship was found between taking an information security course and being attacked ($p = 0.199$, $X^2 = 1.647, df = 1$). This suggests that general awareness training alone may not be sufficient to prevent attacks on SNS.

5. Discussion and conclusion

The findings of this study provide important insights into how users' behavioral intentions influence information security on social networking sites. As cyber threats increasingly exploit both human and technological vulnerabilities, proactive user behavior plays a decisive role. Statistical analysis of 151 respondents confirmed significant associations between security incidents and specific habits, including password update frequency (RQ#1), (although contemporary guidelines increasingly discourage frequent password rotation in favor of strong and unique credentials, the observed association may reflect broader security-conscious behavior rather than the isolated effect of password updates alone), the extent of personal information sharing (RQ#2). The finding that respondents who "never" share personal information reported no attacks should be interpreted cautiously, as such users may also exhibit minimal platform engagement or may not perceive certain incidents as security breaches, and the active use of privacy settings (RQ#3). These results are consistent with prior research highlighting how careless data-sharing practices increase exposure to malicious actors (Amuda et al., 2022) and align with current security recommendations, such as disabling location services (NSA, 2025). A key contribution of this study concerns RQ#4, where no significant relationship was found between formal security awareness and attack prevention. In contrast to Herath, Khanna, and Ahmed (2022), the results suggest that among the active working-age population, optimism bias and the desire for social engagement often outweigh theoretical security knowledge. Users may be aware of risks in principle, yet still engage in risky behavior due to a perceived personal invulnerability. Overall, the findings reinforce that information security on SNS is primarily a behavioral challenge rather than a purely technical one. These findings support a shift from awareness-driven to behaviorally embedded security models, aligning with emerging "security-by-design" and "security-by-default" paradigms. While limited by sample size, the study indicates that habitual measures such as credential rotation and strict visibility management provide more effective protection than passive awareness alone. Future research should examine how platform-specific algorithms and interface design can support the development of secure user habits and foster a culture of "security-by-habit".

5.1. Implications for practice

The findings show that information security on social networking sites depends primarily on habitual user behavior rather than formal awareness training alone. Since security education did not significantly reduce attack experience, practitioners should focus on reinforcing secure daily practices through automation, default-secure settings, and context-aware prompts. Platform design and privacy-by-default mechanisms play a key socio-technical role in shaping safer behavior. Overall, embedding security into routine information management practices supports a security-by-habit approach that is more effective than reliance on passive knowledge-based measures.

5.2. Limitations and further research

This study has several limitations. The use of non-probability sampling and self-reported data may reduce generalizability and introduce response bias. In addition, information security incidents were analyzed only at an aggregate level, without distinguishing attack types or severity. The dependent variable relies on self-reported experience, although some breaches (e.g., data leakage) may remain unnoticed, leading to underreporting. Future research should employ larger, more representative samples, adopt longitudinal designs, and further explore socio-technical interactions, particularly how digital environments and interface design shape habitual security-related user behavior.

Acknowledgment

The authors thank Thi Hang Nguyen for her strong research and assistance with the interviews, data collection, dissemination, and statistical analysis. This research was supported by the Internal Grant Agency of the Faculty of Economics and Management, Czech University of Life Sciences Prague, grant no. IGA2025A1012 Harmonization of life sciences data for HPC calculations.

References

Ali, R. F., Dominic, P. D. D., & Ali, K. (2020). Organizational governance, social bonds and information security policy compliance: a perspective towards oil and gas employees. Sustainability, 12(20), 1–27. https://doi.org/10.3390/SU12208576

Ali, R. F., Dominic, P. D. D., Ali, S. E. A., Rehman, M., & Sohail, A. (2021). Information security behavior and information security policy compliance: a systematic literature review for identifying the transformation process from noncompliance to compliance. Applied Sciences, 11(8). https://doi.org/10.3390/APP11083383

DataReportal. (2022). Digital 2022: Global Overview Report - DataReportal – Global Digital Insights. https://datareportal.com/reports/digital-2022-global-overview-report

Egelman, S., & Peer, E. (2015). Scaling the security wall: Developing a security behavior intentions scale (SeBIS). In CHI 2015, Proceedings of the 33rd Annual Conference on Human Factors in Computing Systems (pp. 2873–2882). https://doi.org/10.1145/2702123.2702249

Hadlington, L. (2017). Human factors in cybersecurity; examining the link between Internet addiction, impulsivity, attitudes towards cybersecurity, and risky cybersecurity behaviours. Heliyon, 3(7), e00346. https://doi.org/10.1016/J.HELIYON.2017.E00346

Herath, T. B. G., Khanna, P., & Ahmed, M. (2022). Cybersecurity Practices for Social Media Users: A Systematic Literature Review. Journal of Cybersecurity and Privacy, 2(1), 1–18. https://doi.org/10.3390/JCP2010001

Lin, C. J., & Jia, H. (2023). Time Pressure Affects the Risk Preference and Outcome Evaluation. International Journal of Environmental Research and Public Health, 20(4). https://doi.org/10.3390/IJERPH20043205

Purdue Global. (2023). Top 10 Worst Data Breaches of All Time [Infographic]. https://www.purdueglobal.edu/blog/information-technology/worst-data-breaches-infographic/

Wang, C., Wang, Y., Chen, Y., Liu, H., & Liu, J. (2020). User authentication on mobile devices: Approaches, threats and trends. Computer Networks, 170, 107118. https://doi.org/10.1016/J.COMNET.2020.107118

West, R. (2008). The psychology of security. Communications of the ACM, 51(4), 34–40. https://doi.org/10.1145/1330311.1330320

Whitman, M. E.., & Mattord, H. J. (2019). Principles of information security, 728. https://books.google.com/books/about/Principles_of_Information_Security.html?hl=cs&id=59dUDgAAQBAJ

Wiederhold, B. K. (2014). The role of psychology in enhancing cybersecurity. Cyberpsychology, Behavior, and Social Networking, 17(3), 131–132. https://doi.org/10.1089/CYBER.2014.1502

BETWEEN CREATIVITY AND MANIPULATION: HOW CZECH ONLINE NEWS FRAMES AI-GENERATED CONTENT, PLATFORMS, AND RESPONSIBILITY

Lukas Herout, Anna Ulrichova

Skoda Auto University, Mlada Boleslav
lukas.herout@savs.cz, anna.ulrichova@savs.cz

DOI: 10.35011/IDIMT-2026-205

Keywords

AI-generated content; deepfakes; social platforms; media framing; responsibility; online news.

Abstract

The rise of generative AI has expanded the creation of synthetic content that can imitate a person's image, voice, or identity. On social media and digital platforms, such content is not only a space for creativity but also a source of manipulation, fraud, and identity misuse. This study analyses how Czech online news framed deepfakes and other forms of AI-generated and manipulated content in 2025, with a specific focus on platforms and responsibility. The analysis is based on a quantitative content analysis of 47 articles published by the most visited Czech online news websites. The results show that the topic was most often framed as a practical risk for individuals, with user verification and caution presented as the main response. Platform context was highly visible in the articles, but platform responsibility was addressed much less often. The study therefore shows a tension between a platform-distributed problem and an individualised solution.

1. Introduction

Generative AI has changed the conditions of digital content creation. Images, voices, videos, and even personal identities can now be imitated quickly, cheaply, and convincingly. On social media and digital platforms, this creates a clear tension: the same technologies that expand creativity, personalised communication, and participatory content creation can also be used for manipulation, fraud, identity misuse, and harmful content distribution.

Deepfakes and other forms of AI-generated and manipulated content are therefore not only a technical novelty. Their social meaning also depends on how they spread, how they are explained to the public, and who is seen as responsible for their misuse. Social media and digital platforms play a key role here, as they provide space for distribution, visibility, sharing, targeting, and moderation. This is where the creative potential of AI meets the risk of misuse – and where responsibility may be assigned to users, offenders, institutions, or platforms.

News media do not only describe this tension. They also frame it. They shape whether deepfakes are presented mainly as a technological threat, criminal behaviour, a failure of user caution, a platform problem, a regulatory challenge, or a broader social risk. A framing perspective therefore makes it possible to analyse how cases of AI-manipulated content are organised in news coverage, who is

assigned responsibility, and what solutions are offered to the public (Entman, 1993; de Vreese, 2005; Semetko & Valkenburg, 2000).

This paper analyses how Czech online news framed deepfakes and other forms of AI-generated and manipulated content in 2025, with a specific focus on social platforms and responsibility for misuse. The study is based on a quantitative content analysis of 47 articles published by the most visited Czech online news websites. The research is guided by three questions:

- RQ1: How does Czech online news frame deepfakes and other forms of AI-generated and manipulated content in the context of digital creation, identity misuse, and online distribution?
- RQ2: What role do the analysed texts assign to social media and digital platforms in spreading, enabling, or addressing this content?
- RQ3: How is responsibility distributed in media coverage among offenders, users, platforms, and institutional actors?

2. Theoretical background

2.1. Media framing and responsibility

Media framing helps to show how news coverage selects, highlights, and connects events into an interpretive frame. Frames shape how a problem is defined, which causes are stressed, who is assigned responsibility, and which solutions are presented as appropriate (Entman, 1993). This approach is useful for analysing deepfakes and other forms of AI-generated and manipulated content because it links the description of the problem with blame, prevention, and possible responses.

Media content may include both general frames, such as responsibility, human interest, or economic consequences, and issue-specific frames adapted to a concrete topic (Semetko & Valkenburg, 2000; de Vreese, 2005). For AI-generated content, it is important whether the issue is presented as an individual problem, a technological threat, criminal behaviour, institutional failure, a platform problem, or a broader social risk. Episodic victim stories may support an individualised view of the problem, while broader thematic context can open space for systemic and platform responsibility (Boukes, 2022).

2.2. Deepfakes as harmful digital creation

Deepfakes and other forms of synthetic content matter not only because they can create convincing false images, videos, or audio recordings, but also because they weaken trust in digital content. Vaccari and Chadwick (2020) show that synthetic political videos may not necessarily change public beliefs on a large scale, but they can increase uncertainty and reduce trust in news. Pawelec (2022) also notes that deepfakes may threaten public deliberation, accountability, and the legitimacy of decision-making.

The issue is not limited to political disinformation. AI-generated and manipulated content can be used for fraud, fake advertising, voice cloning, identity misuse, and non-consensual sexualised content. Gosse and Burkell (2020) show that media representations of deepfakes include several problem definitions – from easy creation and distribution to political threats and sexualised identity abuse. Brooks (2021) adds that the way the problem is framed is linked to the solutions considered relevant: technical framing points to detection, while social or ethical framing opens space for education, regulation, and actor responsibility.

2.3. Platforms between distribution, moderation, and responsibility

For social media, creativity, and innovation, a key point is that AI-generated content does not operate in isolation. Its meaning is shaped by the environment in which it is posted, shared, recommended, monetised, or moderated. Social media and digital platforms are therefore not only passive channels of distribution. They may serve as spaces of circulation, infrastructures that enable reach and targeting, and actors expected to moderate, label, detect, or remove harmful content.

Recent literature shows that the platform dimension is increasingly important for synthetic content. Cochran and Napshin (2021) focus on awareness of deepfakes, public concerns, and online platform accountability. Corsi et al. (2024) show on the case of X that synthetic media are becoming a more common part of the social media environment, even if not all uses are harmful. Fisher et al. (2024) argue that responses to synthetic content should not be reduced to bans or technical detection, but should be integrated into broader content moderation rules. One concrete response is the labelling of AI-generated content, where label design affects how users interpret content on social media (Gamage et al., 2025).

Technology labels and distribution context also matter. Rauchfleisch, Vogler, and de Seta (2025) show that the term deepfakes is more strongly linked to risks, crime, and pornography than the more neutral term synthetic media. Bode (2021) also notes that reception is shaped not only by the content itself but also by its platform context and presentation.

This study therefore focuses on the tension between three layers: AI as a tool of digital creation, platforms as spaces of distribution, and responsibility for misuse. This makes it possible to examine whether Czech online news presents deepfakes mainly as a matter of individual caution, or whether it also opens the issue of platform, institutional, and regulatory responsibility.

3. Methodology

3.1. Research design and corpus

The study was conducted as a quantitative content analysis of Czech online news. The media frame included the fifteen most visited Czech online news websites according to NetMonitor data for 2025. The analysed period covered 1 January to 31 December 2025.

Articles were searched through Google queries using the site: operator and keywords related to deepfake videos, voice cloning, vishing, AI scams, fake videos, fake voices, impersonation, and fake advertising. For the most represented outlet, Novinky.cz, the search strategy was also checked in the NewtonOne database.

After deduplication, 665 items were assessed. In total, 82 texts entered detailed screening, and 47 articles were included in the final corpus after full-text relevance assessment. Only texts in which deepfakes or another form of AI-generated and manipulated content formed a central or clearly relevant topic were included.

The search strategy was designed to be broad and transparent, but it was still dependent on keyword selection, search engine indexing, and the availability of full texts on the original websites. Some relevant articles may therefore have remained outside the corpus, especially if they used unusual terminology, were no longer indexed, or were not fully accessible. The corpus should therefore be understood as a systematically constructed set of relevant texts rather than as a complete archive of all possible mentions.

3.2. Coding scheme

The unit of analysis was one online news article. The coding scheme included the type of case, AI content modality, primary and secondary media frame, responsibility attribution, communicated solutions, and selected indicators of prevention, institutional response, technical detection, and victim story presence.

Given the focus of the paper, three platform-oriented variables were also coded. platform_context_present captured whether the article mentioned social media, digital platforms, messaging apps, fake websites, or other online environments as spaces of distribution, contact, publication, removal, or regulation. platform_role distinguished whether platforms appeared as a neutral distribution environment, an enabling infrastructure, an actor of solution, or only a marginal element. platform_accountability_present captured whether the text addressed platform responsibility, including the duty to intervene, cooperate, moderate, label, or remove problematic content.

The variables platform_role and platform_accountability_present were treated as analytically separate. platform_role captured how platforms functioned in the article narrative, while platform_accountability_present captured whether the text explicitly addressed platform responsibility, duties, or expected action. A platform could therefore appear as an actor of response without the article clearly assigning systemic accountability to it.

3.3. Coding and analytical procedure

Coding was conducted by two coders using a shared codebook. Key variables were coded independently, and disagreements were resolved by consensus. Intercoder reliability was tested using Cohen's kappa. Agreement was high for the main content variables, including primary_frame (κ = 0.973), blame_actor_primary (κ = 0.943), and treatment_primary (κ = 0.935). For the platform variables, kappa reached κ = 0.644 for platform_context_present, κ = 0.818 for platform_role, and κ = 0.873 for platform_accountability_present.

The analysis was descriptive. It first examined dominant frames and communicated solutions. It then focused on the role of social media and digital platforms and on how responsibility was distributed among offenders, users, platforms, and institutional actors. Differences between media outlets were interpreted only as indicative because of the uneven distribution of articles across outlets.

4. Results

The final corpus included 47 articles. Coverage was uneven across media outlets: Novinky.cz accounted for the largest share (20 articles; 42.6%), followed by Seznam Zprávy and iROZHLAS with five articles each. Due to this imbalance, differences between outlets were interpreted only as indicative. Novinky.cz more often used the victim frame, while Seznam Zprávy and iROZHLAS placed more emphasis on prevention, institutional response, and broader social context.

4.1. Framing of AI-generated and manipulated content

The most frequent primary frame was *prevention / education* (14 articles; 29.8%), followed by *victim / human interest* (10 articles; 21.3%). Other relevant frames were *institutional response / regulation* (8 articles; 17.0%), *technological escalation / detection* (7 articles; 14.9%), and *broader social threat / erosion of trust* (7 articles; 14.9%). *Responsibility* appeared as the primary frame in only one article.

Deepfakes were therefore not framed mainly as a technological novelty, but rather as a practical risk linked to individual protection, institutions, and trust in the digital environment.

4.2. Social platforms as spaces of distribution and response

Platform or social media context was present in 43 of 47 articles (91.5%). Most texts mentioned social media, digital platforms, messaging apps, fake websites, online advertising, or another environment in which AI-generated content circulated, gained visibility, or reached victims.

As shown in Table 1, platforms were not framed in one single way. Most often, they appeared as a neutral space of *distribution or contact* (18 articles; 38.3%). In 17 articles (36.2%), they were framed as *actors of response / moderation / removal / detection*, for example through moderation, blocking, content removal, AI-content labelling, or cooperation in verifying authenticity. In seven articles (14.9%), platforms were framed as *enabling or co-responsible infrastructure*, meaning that their advertising, data, or distribution mechanisms made circulation or misuse easier.

Table 1. Role of platforms in the corpus

Platform role	Number of articles	Share (%)
Platform not mentioned	4	8.5
Neutral space of distribution / contact	18	38.3
Enabling or co-responsible infrastructure	7	14.9
Actor of response / moderation / removal / detection	17	36.2
Marginal or unclear role	1	2.1
Total	47	100.0

Source: authors

The key contrast is between the visibility of platform context and the much weaker visibility of platform accountability. Although platform context was present in almost all articles, platform accountability was addressed in only 10 articles (21.3%). Platforms were therefore visible mainly as spaces of distribution or possible intervention, but less often as actors carrying systemic responsibility.

4.3. Responsibility and communicated solutions

Responsibility was attributed mainly to offenders, fraudsters, and attackers. This category appeared as the primary responsibility actor in 36 articles (76.6%). Platforms or social media were identified as the main responsibility actor in only three articles (6.4%), and the state or regulators in one article (2.1%).

The same individualising tendency appeared in communicated solutions. The most frequent primary solution was *individual verification / caution* (27 articles; 57.4%). Readers were advised to verify identity, avoid time pressure, contact the source through another channel, or critically assess suspicious video, voice, or advertising content. Technical detection, institutional intervention, and regulation appeared as the main solution in only four articles each.

The main result is therefore the tension between platform-based distribution and individualised response. AI-generated and manipulated content was often located in the environment of social media and digital platforms, but the practical response was most often directed at users. The problem was platform-distributed, while the solution was mostly individualised. This pattern is summarised in Figure 1: AI-generated content is often placed in a platform environment, but the practical response is mainly framed as individual caution.

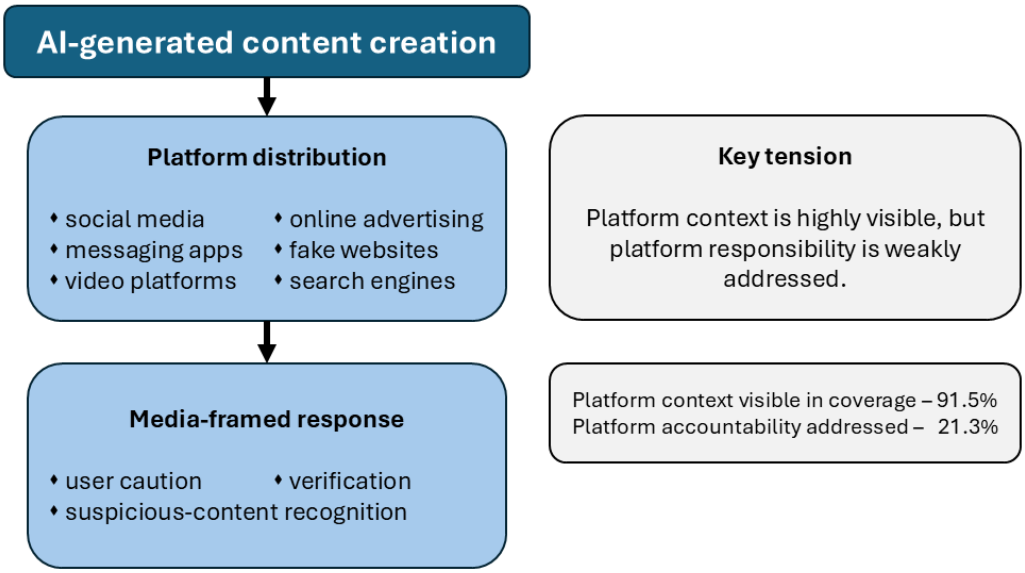


Figure 1. Platform-distributed problem and individualised solution in media framing of AI-generated content
Source: authors

5. Discussion

The results show a clear pattern: Czech online news frames deepfakes and other forms of AI-generated and manipulated content mainly as a practical risk for individuals. The dominant frame is *prevention / education*, and the most common solution is *individual verification / caution*. Readers are therefore placed in the role of those who should be careful, verify sources, avoid pressure, and recognise suspicious videos, voices, or advertisements.

This finding follows earlier research showing that deepfakes and synthetic media are usually linked to risk, deception, misuse, and declining trust (Gosse & Burkell, 2020; Vaccari & Chadwick, 2020; Rauchfleisch et al., 2025). The Czech case follows this risk-oriented pattern, but it also shows a more practical and everyday form of coverage. Compared with studies that emphasise political deepfakes, democratic disruption, or general information uncertainty, the Czech corpus gives strong visibility to fake advertisements, investment scams, voice cloning, identity misuse, and sexualised deepfake content.

The key finding concerns the link between platforms and responsibility. Platform context was very common in the analysed texts, but platforms were less often framed as responsible actors. Social media, messaging apps, fake websites, online advertising, and video platforms appeared as spaces where AI content circulates, gains visibility, or reaches victims. Yet blame was mostly assigned to offenders, while practical defence was shifted to users.

This is the central tension of the paper. AI-generated content is enabled by new creative technologies, and its reach is strengthened by platform environments. However, media coverage more often leads to appeals for individual caution than to a debate about platforms, advertising systems, moderation, content labelling, or responsibility. In short, the problem is platform-distributed, but the solution is individualised.

This finding is also relevant in the European regulatory context. The Digital Services Act strengthens platform duties related to illegal content reporting, advertising transparency, content moderation, and the mitigation of systemic risks. The AI Act develops a broader risk-based framework for trustworthy

AI. From this perspective, the weak visibility of platform accountability in Czech news coverage is important: media framing still tends to place practical responsibility on users, even though European regulation increasingly treats platforms and AI providers as actors with their own duties (European Commission, 2026a; European Commission, 2026b).

This finding has practical implications for debates on social media and AI-enabled creativity. If harmful synthetic content is framed mainly as a matter of user caution, then the burden of response is shifted to individuals who often have limited tools for verification. A stronger platform-oriented framing would make visible not only malicious actors, but also the role of advertising systems, recommender mechanisms, content moderation, and labelling practices. This does not mean that users have no responsibility, but it suggests that individual caution alone is not a sufficient response to platform-distributed AI misuse.

This conclusion is in line with Brooks (2021), who argues that problem definition shapes the type of solution that appears relevant. It also supports the view that synthetic content cannot be addressed only through technical detection. Fisher et al. (2024) argue that responses to generative AI should be integrated into broader content moderation rules, while labelling synthetic content is discussed as one way to help users understand the origin and nature of content on social media (Gamage et al., 2025). In the Czech coverage analysed here, however, technical detection, platform moderation, and labelling remained secondary to a warning logic directed at the reader.

Overall, news coverage recognises the risks of platform-distributed AI content, but addresses platform responsibility only to a limited extent. This shows the ambivalent position of AI-generated content in social media environments: it is linked to creativity and innovation, but also to manipulation, misuse, and unresolved questions of responsibility.

The study has several limitations. The corpus includes only Czech online news articles from 2025, and the number of relevant texts is relatively small. The results should therefore be interpreted as a descriptive view of media framing rather than as a basis for strong generalisation across all media types. The analysis also focuses on news representation, not on actual platform policies or user behaviour.

6. Conclusion

This paper examined how Czech online news frames deepfakes and other forms of AI-generated and manipulated content in relation to social platforms and responsibility for misuse. The analysis of 47 articles from 2025 showed that the topic was most often presented as a practical risk against which users should protect themselves through caution, verification, and the ability to recognise suspicious content.

The main finding is the tension between platform-based distribution and individualised response. Social media, digital platforms, online advertising, messaging apps, and fake websites were highly visible in the coverage, but platform responsibility was addressed less often. Blame was most often assigned to offenders, while the main defence was directed at individual users. Deepfakes were therefore framed not only as a technological phenomenon, but as a social problem linked to trust, security, identity, and responsibility. Future research could compare this pattern across countries and further examine when platforms are presented only as spaces of distribution and when they are framed as responsible actors.

Acknowledgement

This work was supported in part by the Long-Term Conceptual Development of the Research Organization (2026) at Skoda Auto University, Czech Republic.

References

Bode, L. (2021). Deepfaking Keanu: YouTube deepfakes, platform visual effects, and the complexity of reception. Convergence: The International Journal of Research into New Media Technologies, 27(4), 1037–1053. https://doi.org/10.1177/13548565211030454

Boukes, M. (2022). Episodic and thematic framing effects on the attribution of responsibility: The effects of personalized and contextualized news on perceptions of individual and political responsibility for causing the economic crisis. The International Journal of Press/Politics, 27(2), 374–395. https://doi.org/10.1177/1940161220985241

Brooks, C. F. (2021). Popular discourse around deepfakes and the interdisciplinary challenge of fake video distribution. Cyberpsychology, Behavior, and Social Networking, 24(3), 159–163. https://doi.org/10.1089/cyber.2020.0183

Cochran, J., & Napshin, S. (2021). Deepfakes: Awareness, concerns, and platform accountability. Cyberpsychology, Behavior, and Social Networking, 24(3), 164–172. https://doi.org/10.1089/cyber.2020.0100

Corsi, G., Marino, B., & Wong, W. (2024). The spread of synthetic media on X. Harvard Kennedy School Misinformation Review. https://doi.org/10.37016/mr-2020-140

de Vreese, C. H. (2005). News framing: Theory and typology. Information Design Journal + Document Design, 13(1), 51–62. https://doi.org/10.1075/idjdd.13.1.06vre

Entman, R. M. (1993). Framing: Toward clarification of a fractured paradigm. Journal of Communication, 43(4), 51–58. https://doi.org/10.1111/j.1460-2466.1993.tb01304.x

European Commission. (2026a). The Digital Services Act. Shaping Europe's Digital Future. https://digital-strategy.ec.europa.eu/en/policies/digital-services-act

European Commission. (2026b). AI Act. Shaping Europe's Digital Future. https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai

Fisher, S. A., Howard, J. W., & Kira, B. (2024). Moderating synthetic content: The challenge of generative AI. Philosophy & Technology, 37(4). https://doi.org/10.1007/s13347-024-00818-9

Gamage, D., Sewwandi, D., & Zhang, M. (2025). Labeling synthetic content: User perceptions of label designs for AI-generated content on social media. Proceedings of the CHI Conference on Human Factors in Computing Systems, 1–29. https://doi.org/10.1145/3706598.3713171

Gosse, C., & Burkell, J. (2020). Politics and porn: How news media characterizes problems presented by deepfake. Critical Studies in Media Communication, 37(5), 497–511. https://doi.org/10.1080/15295036.2020.1832697

Pawelec, M. (2022). Deepfakes and democracy (theory): How synthetic audio-visual media for disinformation and hate speech threaten core democratic functions. Digital Society, 1, 19. https://doi.org/10.1007/s44206-022-00010-6

Rauchfleisch, A., Vogler, D., & de Seta, G. (2025). Deepfakes or synthetic media? The effect of euphemisms for labeling technology on risk and benefit perceptions. Social Media + Society, 11(3), 1–12. https://doi.org/10.1177/20563051251350975

Semetko, H. A., & Valkenburg, P. M. (2000). Framing European politics: A content analysis of press and television news. Journal of Communication, 50(2), 93–109. https://doi.org/10.1111/j.1460-2466.2000.tb02843.x

Vaccari, C., & Chadwick, A. (2020). Deepfakes and disinformation: Exploring the impact of synthetic political video on deception, uncertainty, and trust in news. Social Media + Society, 6(1), 1–13.

FROM PARTICIPATORY CREATIVITY TO CO-INNOVATION: THE ROLE OF SOCIAL MEDIA AND AI IN INNOVATION ECOSYSTEMS

Lucie Rožmberská

Prague University of Economics and Business
Faculty of Informatics and Statistics
lucie.rozmberska@vse.cz

DOI: 10.35011/IDIMT-2026-213

Keywords

Artificial Intelligence; social media; participatory creativity; co-innovation; digital ecosystems.

Abstract

Social media and generative artificial intelligence (AI) are fundamentally transforming the ways in which innovations emerge in digital environments. However, existing research primarily addresses these phenomena separately and does not provide an explanation of their combined effects within digital innovation ecosystems. This paper responds to this gap and aims to synthesize existing knowledge on the role of social media as a catalyst for innovation and on the impact of AI on participatory creativity, and on this basis proposes a conceptual framework. The paper is based on a review of scholarly literature focusing on social media, participatory creativity, digital communities, AI, and co-innovation. The results indicate three key mechanisms: (1) expansion of the participation base and diversity of inputs, (2) augmentation of creativity through AI, and (3) collective validation and iteration of ideas within digital communities. These mechanisms are integrated into a transformation logic in which user-generated content, through sharing, remixing, and community feedback, can evolve into co-innovations as practically applicable solutions. The main contribution of the paper is the proposal of a conceptual model. In addition, the study contributes to the integration of previously fragmented research and offers implications for both theory and practice of digital innovation.

1. Introduction

Digitalisation fundamentally transforms innovation processes and changes the ways in which organisations, users, and communities create, share, and develop new ideas (Verhoef et al., 2021; Nambisan et al., 2019). In this context, social media have become significant platforms for the generation, dissemination, and collective evaluation of innovations (Kumar et al., 2023). At the same time, generative AI is reshaping creative processes, expanding possibilities for experimentation, and lowering barriers to content creation (Dwivedi et al., 2023; Brynjolfsson et al., 2023). As a result, there is an increasing shift from participatory creativity towards its transformation into collectively validated and organisationally applicable solutions, i.e. co-innovations. Digital communities, platform algorithms, and AI tools play a key role in this process.

Despite this, current research in this area remains fragmented. Individual streams of literature focus on social media, AI, participatory creativity, digital communities, or innovation. However, these phenomena are examined largely in isolation. Consequently, there is a lack of an integrated explanation of how the transition from participatory creativity to organisationally applicable co-innovations occur within digital ecosystems. Existing research therefore only partially captures the interplay of technological, social, and process mechanisms shaping this transition.

The following Table 1 presents the author's analytical synthesis based on a comparison of dominant thematic emphases across different research streams.

Table 1. Author's thematic overview of selected literature and identified gaps

Authors	Research focus	Limitations in relation to this paper
Appel et al. (2020); Kumar et al. (2023)	Social media and marketing / engagement	Limited connection to AI and co-innovation
Dwivedi et al. (2023; 2021); Brynjolfsson et al. (2023)	AI and creativity / digital transformation	Limited consideration of the role of communities and social media
Nambisan et al. (2019); Verhoef et al. (2021)	Digital ecosystems and innovation	Less developed link to participatory creativity and generative AI
Highfield (2021); Abidin (2021)	Participatory culture, memes	Limited connection to organisational and innovation outputs
Sudzina & Stocchetti (2016)	User behaviour on social media	No explicit connection to AI and co-innovation

Source: author

This paper responds to these gaps by integrating these research streams into a unified conceptual framework (see Chapter 5). The contribution of the paper is threefold: (1) it integrates previously separate research streams (social media, AI, participatory creativity), (2) it identifies mechanisms of transformation from content to co-innovation, and (3) it proposes a conceptual model explaining this process.

The aim of the paper is to synthesise insights from the literature review and propose an integrated model.
The research question of the paper is: How does the combination of social media and AI contribute to the transformation of participatory creativity into co-innovations in digital ecosystems?

2. Theoretical background

2.1. Social media and Co-Innovation

Social media (e.g. TikTok, Reddit, Instagram) support open innovation and user involvement in value creation (Nambisan et al., 2019; Kumar et al., 2023). They enable rapid sharing of ideas and their collective evaluation through interactions.

Participatory creativity refers to a process in which users collectively create and modify content, with creativity emerging iteratively within communities (Highfield, 2021).

Co-innovations represent innovations created through collaboration between organisations, users, and communities in digital ecosystems (Nambisan et al., 2019).

2.2. Digital communities as ecosystems

Digital communities can be understood as innovation-oriented ecosystems based on knowledge sharing and collaboration (Verhoef et al., 2021; Dwivedi et al., 2021). They involve interconnected actors, including users, organisations, and platforms, whose interactions contribute to the creation and development of innovations (Adner, 2017).

- AI and Creativity

Generative AI expands creative capacities and accelerates ideation and testing of ideas (Brynjolfsson et al., 2023; Dwivedi et al., 2023). At the same time, it raises questions regarding quality and dependence on algorithms.

- Participatory Culture

Participatory culture is characterised by active user involvement in content creation and sharing. Memes and short-form content support rapid diffusion and adaptation of ideas (Highfield, 2021; Abidin, 2021).

3. Methodology

The paper is based on a literature review aimed at connecting research on social media, AI, participatory creativity, and co-innovation. The approach follows a theory-building intention and is based on a targeted selection of thematically relevant sources.

Sources were selected according to their relevance to the research question and consist mainly of peer-reviewed articles, supplemented by selected monographs and methodological texts that serve to anchor key concepts and constructs.

The chosen approach has certain limitations. It does not represent a fully systematic database review with a reproducible search protocol, but rather a conceptually oriented literature review.

4. Results

The synthesis of relevant sources reveals consistent patterns across the literature. Based on these, three mutually complementary mechanisms explaining the emergence of co-innovations are identified: (1) expansion of the participation base and diversity of inputs, (2) enhancement of creativity through AI tools, and (3) collective validation and iteration leading to co-innovation.

4.1. Social media as a catalyst for Co-Innovation

Existing literature suggests that social media expand the range of actors involved in innovation processes, thereby increasing the diversity of ideas and accelerating their diffusion. This effect is associated with lower entry barriers, a high degree of interactivity, and the possibility of rapid public feedback (Appel et al., 2020).

Illustrative example: The virality of the so-called “TikTok feta pasta” in Finland led to a significant increase in demand for specific ingredients and inspired brands and retailers to adjust their offerings and communication. This case demonstrates how user-generated content can, through community feedback, directly influence market behaviour and lead to practically applicable innovations (Abidin, 2021).

4.2. AI as a creativity amplifier

Another recurring theme in the literature is the role of AI in supporting creative exploration. This role should not be understood as a claim that generating many variants automatically equals creativity. Quantity alone does not make an output creative. In this paper, creativity is understood as a process involving variation, selection, interpretation, contextual judgement, and refinement by human actors and communities. (Dwivedi et al., 2023; Brynjolfsson et al., 2023).

Generative AI can expand the range of possible ideas, combinations, and prototypes. It supports exploration, recombination, and early-stage experimentation by helping users generate alternative formulations, visual concepts, scenarios, or preliminary solutions. In this sense, AI may widen the solution space and accelerate iteration.

However, the creative value of AI-supported outputs depends on human and social evaluation. AI-generated variants become meaningful only when they are interpreted, selected, modified, and validated within a specific community or organisational context. Therefore, AI does not replace human creativity or guarantee innovative outcomes; its amplifying role is conditional.

Illustrative example: Content creators may use generative AI to produce alternative slogans, visuals, product ideas, or prototype descriptions. These outputs are not creative merely because they are numerous or polished. Their potential emerges when users compare them, adapt them to context, test audience reactions, and refine the most promising variants through feedback.

4.3. Communities as innovation environments

Digital communities can be understood as environments in which ideas are iterated, combined, and transferred. Existing studies suggest that interactions between users can lead to the gradual improvement and adaptation of solutions, which represents a significant mechanism of collective creativity (Kumar et al., 2023; Abidin, 2021). This process also includes remixing and sharing of content, whereby users further modify, combine, and transfer existing ideas across communities and platforms, thereby increasing their innovation potential.

Illustrative example: Reddit communities function as crowdsourcing environments where users share ideas, refine solutions, and jointly develop new variants of digital tools and practices. This process illustrates knowledge sharing, iteration, and collective validation typical of co-innovation in digital communities (Dwivedi et al., 2023; Brynjolfsson et al., 2023; Kumar et al., 2023).

4.4. Transformation process

Based on the synthesis, a transformation process linking the above mechanisms can be defined. Within this framework, user-generated content can gradually evolve into co-innovations through participatory creativity, remixing, and community validation. This process is non-linear, iterative, and dependent on the dynamics of digital platforms and community interaction (Muninger et al., 2022).

However, this transformation is not automatic or fully predictable. The model does not imply that every user-generated idea becomes a co-innovation. Success depends on contingent factors such as content visibility, algorithmic amplification, community engagement, feedback quality, organisational responsiveness, and practical applicability. The model therefore explains enabling mechanisms rather than predicts innovation success.

The success rate of the process may differ across platforms, communities, and organisational contexts. Many ideas remain temporary trends or discussion topics, while only some are validated,

further developed, and transformed into applicable solutions. This underlines the selective and uncertain character of co-innovation in digital ecosystems.

This transformation logic provides the basis for the proposal of a conceptual model of AI-enhanced participatory co-innovation, which is presented in Chapter 5.

5. Conceptual model

This chapter systematically translates the results of the synthesis (Chapter 4) into a conceptual model (see Fig. 1). The model is derived directly from the three identified mechanisms: (1) participation, (2) AI-based augmentation of creativity, and (3) collective validation. The following procedure summarises the methodology of its construction and explains how the model operates.

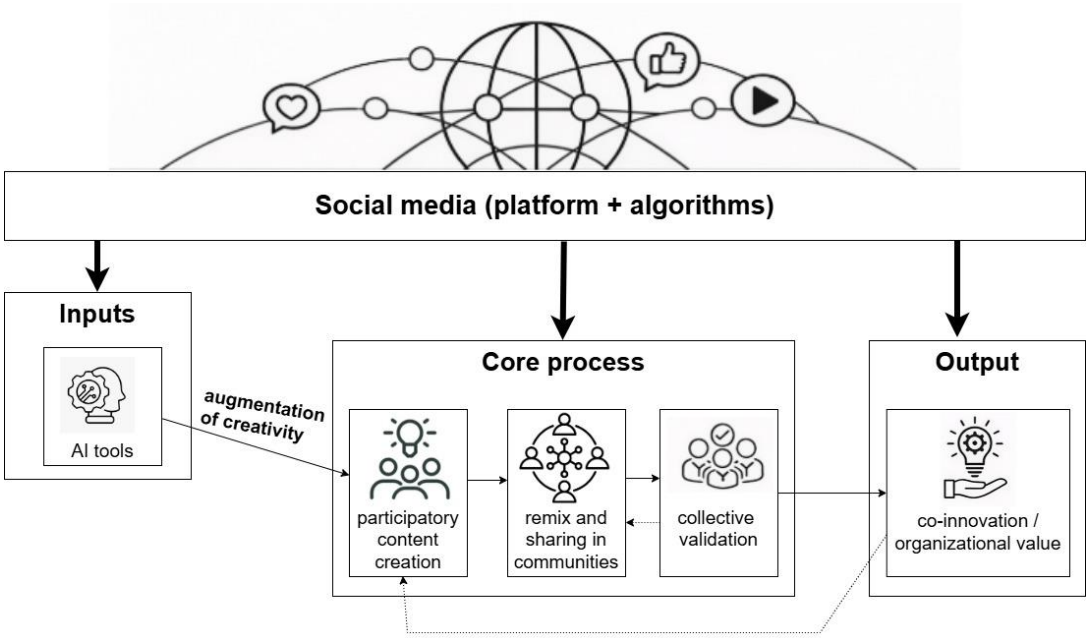


Figure 1. Conceptual model of AI-enhanced participatory co-innovation in digital ecosystems Source: author

5.1. Model development methodology

The model is based on the three identified mechanisms (participation, AI, validation) and translates them into a process scheme. Social media form the context, AI acts as an amplifying factor, and the process phases (creation – remix – validation) lead to the emergence of co-innovations.

5.2. How the model works

The model describes a sequence of five steps: (1) AI supports idea generation, (2) users create content, (3) content is shared and modified, (4) the community provides feedback, and (5) selected ideas may be adopted, adapted, or translated into practical use.

The fifth phase does not imply automatic implementation or universal transformation into co-innovation. It refers to a selective transition from digital circulation to practical application, which occurs only when an idea gains sufficient relevance, adaptability, community support, organisational interest, or market potential.

Many forms of participatory creativity, including memes or platform-specific trends, remain limited to their original community. In such cases, the process may stop at remixing, discussion, or local

validation. The model therefore explains under what conditions some ideas may move beyond their initial digital context, rather than predicting that all digital artefacts become innovations.

5.3. Explanatory and practical use of the model

The proposed model explains how participatory creativity becomes co-innovation in digital ecosystems. It captures the transition from user-generated content to applicable innovation outcomes through three mechanisms: broad participation, AI-supported creative augmentation, and community validation.

The model can be used as an analytical framework for identifying where ideas emerge, how they are modified, and how they gain relevance through collective feedback. For researchers, it links social media, AI, and co-innovation into one explanatory logic. For organisations, it helps locate intervention points in the innovation process: stimulating user participation, using AI to generate and refine variants, evaluating community responses, and selecting ideas with practical potential.

Its practical relevance is strongest in contexts such as digital marketing, community management, customer-driven innovation, early-stage product development, and platform-based entrepreneurship, where innovation depends on the interaction between users, technologies, and organisations.

5.4. Limitations of the model and implications for future research

The proposed model is based on a synthesis of literature and has not been empirically tested. It therefore represents a simplified and generalised view of the transformation from participatory creativity to co-innovation. The model primarily applies to open digital environments with a high degree of user participation and may be less applicable in closed or highly regulated innovation contexts.

The model also has several specific limitations. It only partially captures the role of platforms and algorithms as gatekeepers influencing visibility, diffusion, and selection of ideas. It also does not fully address differences in output quality or potential negative impacts of AI, such as idea homogenisation, dependence on automated tools, or reduced originality.

Finally, the model does not specify the probability or success rate of transformation into co-innovation. It should therefore be understood as an explanatory framework of possible mechanisms rather than as a predictive model. Future research should empirically test the model across different platforms, communities, and organisational settings, and further examine the relationship between human and AI-supported creativity.

6. Discussion

This paper represents a theory-building contribution that integrates previously separate research streams and interprets their relationships in the context of digital innovation processes. The main contribution lies in proposing a coherent model that explains how user-generated content can, through community interactions and technological support, be transformed into practically applicable solutions.

The research question of the paper – how the combination of social media and AI contributes to the transformation of participatory creativity into co-innovations – is answered by identifying three interconnected mechanisms: participation, AI-supported creative exploration, and collective validation. Social media expand the participation base and enable rapid sharing and diffusion of ideas, AI supports the generation and refinement of solution variants, and digital communities provide

feedback and validation. Together, these mechanisms form a process through which selected user-generated ideas may gradually develop into applicable innovation outcomes.

The results suggest that co-innovation in digital environments does not emerge as a linear or automatic process, but rather as the outcome of continuous interactions between users, communities, organisations, and technologies (Nambisan et al., 2017). AI can be understood here as a supporting infrastructure that expands possibilities for ideation, experimentation, and testing, while the relevance and applicability of outputs depend on human interpretation and community evaluation.

At the same time, this development introduces several tensions, for example between creativity and algorithmic optimisation, participation and platform control, or between support and dependence on AI (Dwivedi et al., 2023; Appel et al., 2020; Sudzina & Stocchetti, 2016).

In addition to its explanatory role, the model also has practical relevance. It can serve as a diagnostic framework for organisations seeking to understand how ideas generated on social media platforms can be supported, selected, and further developed with the help of AI and community feedback. This is particularly relevant for organisations working with online communities, digital marketing campaigns, customer-driven innovation, early-stage product development, and platform-based innovation. The proposed model may also be applicable across different digital contexts, although its relevance can vary depending on platform type, community dynamics, and regional specificities (Pavlíček, 2021; Novák, 2020).

7. Conclusion

This paper demonstrates that the integration of social media and generative AI can strengthen participatory and community-driven forms of innovation. Based on an integrative synthesis of the literature, it proposes a conceptual model of AI-enhanced participatory co-innovation in digital ecosystems. The model explains how selected user-generated ideas may, through community interaction, AI-supported exploration, and collective validation, be translated into practically applicable innovation outcomes.

The main contribution of the study lies in integrating previously fragmented research and linking technological, social, and process dimensions of innovation into a single explanatory framework. The model is useful not only as a theoretical explanation of AI-enhanced participatory co-innovation, but also as a practical framework for identifying where organisations can support idea generation, community validation, and the further development of digital content with innovation potential.

At the same time, the study is limited by its conceptual character and the absence of empirical validation. Future research should therefore focus on empirical testing of the model and its application across different types of digital platforms, communities, and organisational contexts.

References

Abidin, C. (2021). Mapping internet celebrity on TikTok: Exploring attention economies and visibility cultures. Cultural Science Journal, 12(1), 77–103. https://doi.org/10.5334/csci.140

Adner, R. (2017). Ecosystem as structure: An actionable construct for strategy. Journal of Management, 43(1), 39–58. https://doi.org/10.1177/0149206316678451

Appel, G., Grewal, L., Hadi, R., & Stephen, A. T. (2020). The future of social media in marketing. Journal of the Academy of Marketing Science, 48(1), 79–95. https://doi.org/10.1007/s11747-019-00695-1

Dwivedi, Y. K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., Duan, Y., Dwivedi, R., Janssen, M., Kumar, V., Olanrewaju, G.-S., Rana, N. P., Williams, M. D., & Wright, R. (2021). Artificial intelligence (AI):

Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. International Journal of Information Management, 57, 101994. https://doi.org/10.1016/j.ijinfomgt.2019.08.002

Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., Baabdullah, A. M., Koohang, A., Raghavan, V., Ahuja, M., Albanna, H., Albashrawi, M. A., Al-Busaidi, A. S., Balakrishnan, J., Barlette, Y., Basu, S., Bose, I., Brooks, L., Buhalis, D., Cai, X., & Wright, R. (2023). "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. International Journal of Information Management, 71, 102642. https://doi.org/10.1016/j.ijinfomgt.2023.102642

Brynjolfsson, E., Li, D., & Raymond, L. R. (2023). Generative AI at work. NBER Working Paper No. 31161. https://doi.org/10.3386/w31161

Highfield, T. (2021). Memes, GIFs, and remixes: Digital media and participatory cultures. Polity Press.

Kumar, V., Dixit, A., Javalgi, R. G., & Dass, M. (2023). Engaging customers through social media: A review and research agenda. Journal of Business Research, 156, 113432. https://doi.org/10.1016/j.jbusres.2022.113432

Muninger, M.-I., Mahr, D., & Hammedi, W. (2022). Social media use: A review of innovation management practices. Journal of Business Research, 143, 140–156. https://doi.org/10.1016/j.jbusres.2022.01.039

Nambisan, S., Lyytinen, K., Majchrzak, A., & Song, M. (2017). Digital innovation management. MIS Quarterly, 41(1), 223–238. https://doi.org/10.25300/MISQ/2017/41:1.03

Nambisan, S., Wright, M., & Feldman, M. (2019). The digital transformation of innovation and entrepreneurship. Research Policy, 48(8), 103773. https://doi.org/10.1016/j.respol.2019.03.020

Novák, R. A. (2020). Digitální transformace a ko-inovace v organizacích. VŠE Publishing.

Pavlíček, A. (2021). Sociální média a marketingová komunikace. Oeconomica.

Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. Journal of Business Research, 104, 333–339. https://doi.org/10.1016/j.jbusres.2019.07.039

Sudzina, F., & Stocchetti, M. (2016). Understanding the impact of social media on consumer behavior. Journal of Systems and Information Technology, 18(2), 168–185. https://doi.org/10.1108/JSIT-08-2015-0055

Torraco, R. J. (2005). Writing integrative literature reviews: Guidelines and examples. Human Resource Development Review, 4(3), 356–367. https://doi.org/10.1177/1534484305278283

Verhoef, P. C., Kannan, P. K., & Inman, J. J. (2021). Digital transformation: A multidisciplinary reflection and research agenda. Journal of Business Research, 122, 889–901. https://doi.org/10.1016/j.jbusres.2019.09.022

RE-EXAMINING THE LINK BETWEEN CHRONOTYPE AND SOCIAL NETWORKING SITES USE FREQUENCY

František Sudzina, Antonín Pavlíček

Prague University of Economics and Business
Faculty of Informatics and Statistics
frantisek.sudzina@vse.cz, antonin.pavlicek@vse.cz

DOI: 10.35011/IDIMT-2026-221

Keywords

Chronotype; social networking sites; frequency of use.

Abstract

Understanding the drivers of digital engagement on social networking sites (SNS) is crucial for platform design and management. Chronotype, an individual's preference for morning or evening activity, is an established predictor of digital behavior, with existing research linking evening-types to more intensive SNS use. This study challenges this paradigm by examining usage frequency across eleven distinct SNS platforms using a four-category chronotype classification (lark, owl, hummingbird, low sleep requirement). Data were collected from university students via a web-based survey and analyzed using the Kruskal-Wallis test. The main finding was a lack of statistically significant differences in usage frequency for most platforms, including Facebook and Instagram, across the chronotype groups. However, users with a low sleep requirement reported significantly more frequent Twitter use. These results suggest that the link between chronotype and SNS use is more complex than previously understood and is highly dependent on the metrics used. The study concludes that simple frequency of use is an insufficient measure of user engagement and that different platforms may serve distinct needs for users with different temporal dispositions, a critical insight for designing content delivery algorithms and analyzing user behavior on digital platforms.

1. Introduction

Social networking sites (SNS) are deeply integrated into daily life, particularly for young adults. Swain and Pati (2021) highlight that SNS engagement significantly impacts psychosocial behavior, academic performance, and sleep, underscoring the need to identify underlying individual difference factors. One such fundamental factor is chronotype, an individual's inherent biological predisposition for the timing of sleep and activity across a 24-hour period. Horne and Östberg (1976) established the classic typology of morning-types ("larks") and evening-types ("owls").

This trait is highly relevant to university students. Late adolescence and early adulthood feature a pronounced biological shift toward eveningness (Roenneberg, Wirz-Justice, & Merrow, 2003), with approximately one-third of Norwegian university students identifying as definite evening-types (Kjørstad et al., 2022). An emerging body of research has begun to connect this high prevalence of eveningness with patterns of digital media use, suggesting that chronotype may be a key predictor of

how young adults engage with SNS, which is highly relevant for social media operators seeking to optimize content scheduling and user engagement strategies.

However, significant gaps remain in the literature. Much of the existing empirical work has focused on a single platform, most often Facebook, or has examined the relationship through the lens of problematic use and addiction rather than normative usage patterns across a diverse media landscape. Furthermore, chronotype is often operationalized using a simple morning-evening dichotomy, which may obscure more subtle behavioral patterns. For instance, such a classification does not typically differentiate between individuals who are flexible in their sleep-wake habits or those who operate on a chronically restricted sleep schedule.

Therefore, this study was designed to address these gaps. It investigates the differences in usage patterns across a comprehensive array of eleven distinct social networking sites, from Facebook and Instagram to TikTok and Viber. It employs a four-category self-reported chronotype classification - lark, owl, hummingbird, and low sleep requirement - to capture a more granular view of sleep-wake behaviors. The primary aim is to determine whether the established link between eveningness and higher SNS engagement holds across this diverse ecosystem of platforms and to explore how usage patterns differ for individuals with flexible or chronically short sleep schedules.

2. Theoretical background

To understand the relationship between an individual's chronotype and their use of SNS, it is necessary to integrate two distinct but complementary theoretical streams: the psychological motivations for media use and the biological underpinnings of daily human rhythms. The Uses and Gratifications (U&G) theory provides a framework for *why* people actively select media, while the field of chronobiology explains the fundamental individual differences in *when* people are inclined to be active. The synthesis of these perspectives provides a robust foundation for hypothesizing that an individual's temporal disposition will predict their patterns of SNS engagement.

2.1. The active audience: A uses and gratifications perspective on SNS

U&G theory posits that media audiences are active and goal-directed, consciously selecting media to satisfy social and psychological needs (Katz, Blumler, & Gurevitch, 1973). This framework helps explain diverse SNS motivations. In their study applying U&G to Facebook, Papacharissi and Mendelson (2011) identified several key gratifications sought by users, including passing time, relaxing entertainment, companionship, and social interaction. These motivations highlight the role of SNS as a tool for both solitary leisure and social connection.

The modern digital landscape is characterized by its eclecticism, a point emphasized in a comprehensive review by Kuss and Griffiths (2017). Different platforms, from the broad-based community of Facebook to the visual-centric worlds of Instagram and Snapchat, offer unique features and may cater to distinct needs. This diversity suggests that a comprehensive understanding of digital sociality requires examining usage patterns across multiple platforms, rather than focusing on a single service.

2.2. Chronotype, social jetlag, and the rhythms of young adulthood

Chronotype is a genetically anchored biological trait reflecting an individual's innate preference for the timing of sleep and activity, creating a spectrum from morning-types (larks) to evening-types (owls) (Horne & Östberg, 1976). This trait is particularly salient in the university student population, as a pronounced biological shift toward eveningness begins in puberty and peaks in the late teens and

early twenties (Randler, 2011; Roenneberg, Wirz-Justice, & Merrow, 2003), before stabilizing again in adulthood. While age is a macro-level predictor of both chronotype and technology use, examining chronotype within a specific, narrow age cohort (such as university students) isolates the micro-level variance driven by biological timing rather than generational differences.

For many evening-types, the rigid schedules of academic and social life create a conflict with their internal clock. This mismatch between biological and social time has been termed *social jetlag* by Wittmann, Dinich, Merrow, and Roenneberg (2006). Social jetlag forces individuals to wake earlier than their biological clock would dictate, leading to an accumulation of sleep debt during the week. This condition is particularly relevant to the "low sleep requirement" category in this study, which serves as a proxy for a population experiencing chronic sleep restriction. A study by Sampasa-Kanyinga, Hamilton, and Chaput (2018) provides strong support for this connection, demonstrating a dose-response relationship where greater social media use is associated with shorter sleep duration in adolescents. Chronotype is also linked to broader personality structures; for instance, Randler, Schredl, and Göritz (2017) found that morningness correlates with higher conscientiousness, while eveningness is associated with greater openness to experience, providing another layer of individual difference that may influence media choices.

2.3. Synthesizing chronotype and SNS use: A review of the evidence

The convergence of U&G theory and chronobiology provides a compelling rationale for why eveningness would be associated with greater SNS use. For an individual who is biologically alert late at night, when opportunities for face-to-face interaction are reduced, SNS platforms offer a readily available source for satisfying needs for social connection, entertainment, and passing time. The very patterns of SNS communication appear to have their own temporal structure; a study by Diwan et al. (2022) analyzing Facebook Messenger data identified distinct circadian (24-hour) and ultradian (12-hour) rhythms in usage, suggesting a natural alignment between biological rhythms and digital social behavior.

Empirical research supports this link. Foundational studies established a direct association between eveningness and intensive, intrusive Facebook use among university students (Blachnio, Przepiorka, & Díaz-Morales, 2015; Horzum & Demirhan, 2017), with evening-type men particularly prone to Facebook addiction (Horzum, Canan Güngören, & Gür Erdoğan, 2022). This relationship extends to the broader digital ecosystem, as evening chronotype significantly predicts problematic smartphone use and higher screen time (Randjelović et al., 2021; Yılbaş & Karadeniz, 2022).

Recent research highlights complex mediating mechanisms. Eveningness is associated with adverse psychological states; Wallinheimo and Evans (2025) demonstrated that loneliness and anxiety fully mediate the link between eveningness and problematic social media addiction, suggesting dysfunctional coping. Sleep disruption is another mediator: Lin et al. (2021) found poor sleep quality and insomnia mediate the eveningness-problematic SNS use association, while Kortesoja et al. (2023) identified late-night media use as mediating eveningness's negative impact on sleep. Finally, objective mobile phone data confirms evening-active individuals maintain larger, more central social networks, shaping digital interaction structures (Aledavood, Lehmann, & Saramäki, 2018). Collectively, this body of work provides a strong theoretical and empirical basis for investigating the specific ways in which different chronotypes engage with the diverse landscape of modern social networking sites.

Based on the consistent evidence linking eveningness to more intensive and problematic use of digital media, the present study sought to test whether this established relationship generalizes across a broader ecosystem of social networking sites when using frequency of use as the primary metric. Therefore, the following hypotheses were formulated:

- H1: Evening-types (owls) will report a significantly higher frequency of use across the eleven measured SNS platforms compared to morning-types (larks).
- H2: Given the theoretical link between social jetlag and compensatory digital media use, individuals in the "low sleep requirement" category will also report a significantly higher frequency of SNS use compared to chronotypes with regular sleep patterns (larks, owls, and hummingbirds).

3. Methodology

A questionnaire was designed to collect information on usage of SNS and self-reported chronotypes. Data were collected using a web-based survey on 1ka.si during the spring semester of 2022. The sample consisted of 529 respondents (289 men and 240 women). Convenience sampling was used; respondents were university students. Respondents did not receive any monetary nor any other type of incentives. Approximately 95% of the respondents were 24 years old or younger.

For the purpose of this research, following SNS were included in the questionnaire: Facebook, WhatsApp, Instagram, TikTok, Twitter, LinkedIn, V kontakte, Snapchat, Skype, Viber, and Tinder.

Use of these SNS was measured on an ordinal scale where 1 meant online connected, 2 meant 1x daily, 3 meant 1x weekly, 4 meant 1x monthly, 5 meant “I have an account, but I do not use it”, and 6 meant “I do not have an account”. These were used as dependent variables.

With regards to the chronotypes, respondents were given a possibility to choose from the following four possibilities: lark (“I wake up early in the morning, go to bed early in the evening”) – coded as 1, owl (“I stay up late in the evening, take a nap in the morning”) – coded as 2, hummingbird (“switches between lark and owl without any problems”) – coded as 3, and low sleep requirement (“I get 4-6 hours of sleep a day, I usually stay up late and then get up early”) – coded as 4.

Since the use of SNS was measured on the ordinal scale, Kruskal-Wallis test (in IBM SPSS 27) will be used to test whether there is a difference in the level of use of social networking sites among users with different chronotypes, and then among users sleeping for 8 hours (the first three categories) and users who sleep less (the fourth category). A significance level of $\alpha = 0.05$ was established for all statistical tests. Prior to analysis, the assumptions for the Kruskal-Wallis test - specifically, the ordinal nature of the dependent variable, independence of observations, and similarly shaped distributions across groups - were verified.

4. Results

To test whether there are any differences in usage of a particular SNS among users with various chronotypes, Kruskal-Wallis test was used. To give a better understanding of usage of the selected SNS, Table 1 provides also ranks of usage, not only the Kruskal-Wallis test. Due to the coding, a lower rank means more usage, a higher rank means less usage.

Based on the established $\alpha = 0.05$ threshold, there are no statistically significant differences among users with different chronotypes, though there is a marginal difference among users of Twitter at the significance level of 0.1.

Table 1. Usage rank of SNS by chronotype and Kruskal-Wallis test

SNS	Lark	Owl	Hummingbird	Less sleep	Kruskal-Wallis H	df	Asymp. Sig.
Facebook	260.05	251.78	281.18	275.83	5.373	3	0.146
WhatsApp	231.82	274.60	260.42	280.77	4.908	3	0.179
Instagram	264.17	268.64	264.69	248.73	1.097	3	0.778
TikTok	293.15	257.01	266.90	258.65	3.514	3	0.319
Twitter	273.72	261.32	278.68	217.46	6.779	3	0.079
LinkedIn	237.78	262.83	280.15	254.67	5.579	3	0.134
V kontakte	266.72	274.03	253.29	262.60	3.580	3	0.311
Snapchat	281.52	263.86	257.49	277.20	1.686	3	0.640
Skype	292.05	253.03	273.05	256.34	5.768	3	0.123
Viber	275.49	267.68	259.85	256.82	0.969	3	0.809
Tinder	282.95	252.85	268.77	287.07	5.795	3	0.122

Source: author

To test whether there are any differences in usage of a particular SNS among users with different length of sleep, Kruskal-Wallis test was used. Table 2 provides ranks of usage and the Kruskal-Wallis test. Since the groups of users who sleep less is the same as in Table 1, the ranks in this column are the same as before.

Table 2. Usage rank of SNS by length of sleep and Kruskal-Wallis test

SNS	Regular length	Less sleep	Kruskal-Wallis H	df	Asymp. Sig.
Facebook	263.97	275.83	0.331	1	0.565
WhatsApp	263.50	280.77	0.566	1	0.452
Instagram	266.55	248.73	0.952	1	0.329
TikTok	265.60	258.65	0.103	1	0.748
Twitter	269.53	217.46	5.281	1	0.022
LinkedIn	265.98	254.67	0.312	1	0.576
V kontakte	265.23	262.60	0.023	1	0.879
Snapchat	263.84	277.20	0.359	1	0.549
Skype	265.83	256.34	0.226	1	0.635
Viber	265.78	256.82	0.195	1	0.659
Tinder	262.90	287.07	1.759	1	0.185

Source: authors

At the $\alpha = 0.05$ level, a statistically significant difference was found only among users of Twitter ($p = 0.022$), indicating that people who sleep less use Twitter more. No other SNS showed significant differences based on sleep length, not even at the significance level of 0.1.

5. Discussion

The primary objective of this study was to investigate whether differences exist in the usage frequency of eleven popular social networking sites among university students with different self-reported chronotypes. Evaluating the formulated hypotheses, Hypothesis 1 (H1) was rejected, as the Kruskal-Wallis tests indicated no statistically significant differences in the usage frequency of most SNS among the four chronotype categories (lark, owl, hummingbird, and low sleep requirement).

Hypothesis 2 (H2) was only partially supported; while individuals in the "low sleep requirement" category did not show higher usage across most platforms, they did report significantly higher usage frequency exclusively for Twitter. This outcome is particularly surprising given the consistent findings in previous literature. For instance, studies by Horzum and Demirhan (2017) and Blachnio, Przepiorka, and Díaz-Morales (2015) both established a clear association between eveningness and more intensive and intrusive Facebook use. This link has been further corroborated in more recent work, which connects evening chronotype to problematic use of smartphones, the internet, and social media in general (Yılbaş & Karadeniz, 2022; Randjelović et al., 2021; Rigó et al., 2024). The divergence of the present study's findings from this established consensus warrants a careful examination of the methodological and conceptual differences.

These discrepant results may stem from variable measurement. First, chronotype was operationalized using a novel, four-category self-report measure rather than validated instruments like the MEQ (Horne & Östberg, 1976) or CSM (Blachnio et al., 2015; Horzum et al., 2022). While "lark" and "owl" align with traditional typologies, the "low sleep requirement" category conflates chronotype (timing preference) with sleep duration (behavior). This group likely represents individuals experiencing significant social jetlag and chronic sleep debt (Roenneberg, Allebrandt, Merrow, & Vetter, 2012; Wittmann, Dinich, Merrow, & Roenneberg, 2006). Treating this highly eveningness-associated state as a separate category may have diluted the statistical power to detect differences between traditional larks and owls.

Second, SNS use was measured via a six-point ordinal frequency scale. Prior studies often measure intensity, duration, or problematic engagement (Lin et al., 2021; Wallinheimo & Evans, 2025). Evening-types may not log in more frequently, but rather engage in longer, intensive late-night sessions—a behavior mediating poor sleep (Kortesoja et al., 2023). The ordinal scale may lack the sensitivity to capture these key differences in engagement duration and quality, which are central to the U&G framework's emphasis on satisfying needs like passing time (Katz, Blumler, & Gurevitch, 1973; Papacharissi & Mendelson, 2011).

Despite the overall null findings, the results for Twitter present a noteworthy exception. A statistically significant difference was found in Twitter usage between individuals with regular and short sleep durations, with the "low sleep requirement" group using the platform more frequently. This finding suggests that the relationship between sleep patterns and SNS use may be platform-specific. Applying the U&G framework, one could hypothesize that Twitter, with its text-based, information-dense format, serves a different set of needs than more visually-oriented or entertainment-focused platforms. Individuals in the "low sleep requirement" group, who may be experiencing time pressure due to their chronic sleep restriction, might gravitate toward Twitter for efficient information surveillance and to quickly catch up on news and events, a different motivation than the more immersive entertainment sought on other platforms. This aligns with the broader understanding that the digital media landscape is eclectic, with different services catering to distinct gratifications (Kuss & Griffiths, 2017).

Finally, several limitations must be acknowledged. The use of a convenience sample of university students, while common in this area of research, limits generalizability. University students possess a highly specific lifestyle characterized by flexible schedules, a lack of full-time employment, and fewer family obligations, which heavily influences both their sleep patterns and media consumption. The primary limitations, however, are the unvalidated, self-report measures for both chronotype and SNS usage frequency. Future research should aim to replicate this multi-platform investigation using validated chronotype scales and more objective measures of SNS engagement, such as server-side logs, API data, or screen time data captured by smartphone applications, as demonstrated by

Randjelović et al. (2021), or network data from mobile phones, as used by Aledavood, Lehmann, and Saramäki (2018).

6. Conclusion

This study investigated the relationship between four self-reported chronotypes and the frequency of use across eleven social networking sites in a sample of university students. The main finding was that, contrary to a growing body of literature, no statistically significant differences were found in the usage frequency of most SNS based on chronotype.

The primary contribution is methodological, highlighting the critical importance of operationalizing chronotype and SNS use. The discrepancy between these null findings and previous positive associations suggests simple frequency measures insufficiently capture how chronotype influences digital behavior. The significant result for Twitter among short sleepers indicates the sleep-SNS relationship may be platform-specific and tied to particular gratifications.

In conclusion, while the broader connection between eveningness and greater digital media engagement remains well-supported, this study serves as a crucial reminder of the complexity of this relationship. It underscores the need for future research to employ validated, multi-dimensional measures to differentiate between the frequency, duration, and intensity of SNS use, and to explore how these distinct patterns of engagement may vary across the diverse ecosystem of modern social media platforms. For social media operators and digital marketers, these findings suggest that optimizing content delivery and scheduling requires moving beyond simple demographic targeting to incorporate nuanced, platform-specific temporal behaviors and user engagement intensities.

Acknowledgement

During the preparation of this paper, AI-based tools (including Gemini 3.1 Pro) were used exclusively for language editing and stylistic improvements. The authors are fully responsible for all aspects of the work, including study design, data processing, statistical analyses, and interpretation of the results.

References

Aledavood, T., Lehmann, S., & Saramäki, J. (2018). Social network differences of chronotypes identified from mobile phone data. EPJ Data Science, 7(1), 46. https://doi.org/10.1140/epjds/s13688-018-0174-4

Blachnio, A., Przepiorka, A., & Díaz-Morales, J. F. (2015). Facebook use and chronotype: Results of a cross-sectional study. Chronobiology International, 32(9), 1315-1319. https://doi.org/10.3109/07420528.2015.1083998

Diwan, A., Swain, R. K., Minz, S., Parganiha, A., & Pati, A. K. (2022). Ultradian, circadian, and circaseptan rhythms in the patterns of usage of Facebook messenger. Biological Rhythm Research, 53(2), 341-349. https://doi.org/10.1080/09291016.2020.1747741

Horne, J. A., & Ostberg, O. (1976). A self-assessment questionnaire to determine morningness-eveningness in human circadian rhythms. International Journal of Chronobiology, 4(2), 97-110.

Horzum, M. B., & Demirhan, E. (2017). The role of chronotype on Facebook usage aims and attitudes towards Facebook and its features. Computers in Human Behavior, 73, 125-131.https://doi.org/10.1016/j.chb.2017.03.038

Horzum, M. B., Canan Güngören, Ö., & Gür Erdoğan, D. (2022). The influence of chronotype, personality, sex, and sleep duration on Facebook addiction of university students in Turkey. Biological Rhythm Research, 53(7), 1105-1115. https://doi.org/10.1080/09291016.2021.1907508

Katz, E., Blumler, J. G., & Gurevitch, M. (1973). Uses and gratifications research. The Public Opinion Quarterly, 37(4), 509-523. https://doi.org/10.1086/268109

Kjørstad, K., Vedaa, Ø., Pallesen, S., Saxvig, I. W., Hysing, M., & Sivertsen, B. (2022). Circadian preference in young adults: Associations with sleep and mental health outcomes from a national survey of Norwegian university students. Chronobiology International, 39(11), 1465-1474. https://doi.org/10.1080/07420528.2022.2121657

Kortesoja, L., Vainikainen, M. P., Hotulainen, R., & Merikanto, I. (2023). Late-night digital media use in relation to chronotype, sleep and tiredness on school days in adolescence. Journal of Youth and Adolescence, 52(2), 419-433. https://doi.org/10.1007/s10964-022-01703-4

Kuss, D. J., & Griffiths, M. D. (2017). Social networking sites and addiction: Ten lessons learned. International Journal of Environmental Research and Public Health, 14(3), 311. https://doi.org/10.3390/ijerph14030311

Lin, C. Y., Imani, V., Griffiths, M. D., Broström, A., Nygårdh, A., Demetrovics, Z., & Pakpour, A. H. (2021). Temporal associations between morningness/eveningness, problematic social media use, psychological distress and daytime sleepiness: Mediated roles of sleep quality and insomnia among young adults. Journal of Sleep Research, 30(1), e13076. https://doi.org/10.1111/jsr.13076

Papacharissi, Z., & Mendelson, A. (2011). Toward a new (er) sociability: Uses, gratifications and social capital on Facebook. In Media Perspectives for the 21st Century (pp. 212-230). Routledge. https://doi.org/10.4324/9780203834077

Randjelovic, P., Stojiljkovic, N., Radulovic, N., Stojanovic, N., & Ilic, I. (2021). Problematic smartphone use, screen time and chronotype correlations in university students. European Addiction Research, 27(1), 67-74. https://doi.org/10.1159/000506738

Randler, C. (2011). Age and gender differences in morningness–eveningness during adolescence. The Journal of Genetic Psychology, 172(3), 302-308. https://doi.org/10.1080/00221325.2010.535225

Randler, C., Schredl, M., & Göritz, A. S. (2017). Chronotype, sleep behavior, and the big five personality factors. Sage Open, 7(3), 2158244017728321. https://doi.org/10.1177/2158244017728321

Rigó, A., Tóth-Király, I., Magi, A., Eisinger, A., Griffiths, M. D., & Demetrovics, Z. (2024). Morningness-eveningness and problematic online activities. International Journal of Mental Health and Addiction, 22(5), 2766-2786. https://doi.org/10.1007/s11469-023-01017-x

Roenneberg, T., Wirz-Justice, A., & Merrow, M. (2003). Life between clocks: daily temporal patterns of human chronotypes. Journal of Biological Rhythms, 18(1), 80-90. https://doi.org/10.1177/0748730402239679

Roenneberg, T., Allebrandt, K. V., Merrow, M., & Vetter, C. (2012). Social jetlag and obesity. Current Biology, 22(10), 939-943. https://doi.org/10.1016/j.cub.2012.03.038

Sampasa-Kanyinga, H., Hamilton, H. A., & Chaput, J. P. (2018). Use of social media is associated with short sleep duration in a dose–response manner in students aged 11 to 20 years. Acta Paediatrica, 107(4), 694-700. https://doi.org/10.1111/apa.14210

Swain, R. K., & Pati, A. K. (2021). Use of social networking sites (SNSs) and its repercussions on sleep quality, psychosocial behavior, academic performance and circadian rhythm of humans–a brief review. Biological Rhythm Research, 52(8), 1139-1178. https://doi.org/10.1080/09291016.2019.1620487

Wallinheimo, A. S., & Evans, S. L. (2025). Mechanisms that link circadian preference to problematic smartphone and social media use in young adults. PLOS ONE, 20(9), e0331961. https://doi.org/10.1371/journal.pone.0331961

Wittmann, M., Dinich, J., Merrow, M., & Roenneberg, T. (2006). Social jetlag: misalignment of biological and social time. Chronobiology International, 23(1-2), 497-509. https://doi.org/10.1080/07420520500545979

Yılbaş, B., & Karadeniz, P. G. (2022). The relationship between chronotype and impulsivity, attention-deficit disorder, internet, social media, and smartphone addiction. Alpha Psychiatry, 23(4), 203-209. https://doi.org/10.5152/alphapsychiatry.2022.21656

ESTIMATION OF THE TRUTH OF NEWS AND THEIR VERIFICATION (SURVEY)

Jana Syrovátková, Antonín Pavlíček

University of Economics, Prague
Faculty of Informatics and Statistics
jana.syrovatkova@vse.cz, antonin.pavlicek@vse.cz

DOI: 10.35011/IDIMT-2026-229

Keywords

Social networks; disinformation; fact-checking.

Abstract

Although disinformation is a more than 100-year-old issue, with the spread of news on the Internet, the phenomenon of Fake News is a current issue that is shaking up the Internet world. In a survey conducted in the fall of 2024 with almost 300 responses, mostly from university students in the Czech Republic, we focused on the behavior of users on the Internet and primarily on social networks. The main goal of this research was to analyze how much they verify the information they read on social networks and how capable they are of detecting specific disinformation. We also found out what social networks and fact-checking tools they use. We found that almost half of the respondents use some form of fact-checking tools on the Internet. We specifically analyzed the 220 respondents who gave a definite true/false judgement on at least one news item. Tool users correctly identified 88.2 % of the items they judged, compared with 90.7 % for non-users. This is consistent with a slightly lower accuracy among tool users, but the difference was not statistically significant (Welch t-test, p = 0.38).

1. Introduction

Disinformation has been around for over 100 years (Gregor & Mlejnková, 2018), but with the development of social networks it has become more widespread and has been referred to as fake news. It is the deliberate misleading of readers, often with financial or political motivation (Tandoc et al., 2018). This distinguishes it from misinformation, which is often spread in good faith, with the spreader not realizing the error of the information (Stahl, 2006).

A more serious way of deceiving users could be called deepfake (from the words fake and deep learning). Initially, it was the creation of a deepfake user on Reddit, who incorporated famous personalities into films. One of the first videos was a video with Barrack Obama in 2017 (Kietzmann et al., 2020). Videos can synchronize movement and sound, use increasingly better techniques using artificial intelligence, and so it is increasingly complicated for users to recognize whether it is a deepfake or reality.

The basic protection against disinformation is media literacy, which mainly includes knowledge of how the media works, understanding the media from the perspective of influencing society, and the ability to assess and analyze media messages (Gyüre, 2024).

Previous research among predominantly young Czech respondents found high self-reported information literacy and a strong declared tendency to verify online information. However, the practical tasks also revealed remaining susceptibility to cognitive fallacies and misleading social media content. Moreover, respondents tended to consider themselves more resistant to fake news than other people, confirming the third-person effect (Sigmund & Korčák, 2022).

One way how to fight against fake news is through fact-checking. This involves either checking news on special websites or tools or using various browser extensions. One of the first groups operating in this area was the French daily Le Monde, which has been involved in fact-checking since 2009 and is even behind one of the first browser extensions – Decodex (Gregor & Mlejnková, 2018).

Currently, there are a number of services and web extensions, the Verifee program checks the credibility of individual articles using artificial intelligence (Verifee, 2024), NewsGuard offers a desktop version as well as a mobile version (NewsGuard Technologies, Inc., n.d.). Google has a whole group of tools that it simultaneously implements into search results (Google News Initiative, n.d.).

The basic research question of the entire study is: "Do users of fact-checking tools differ from non-users in their ability to recognize true and false news?"

2. Survey

The research is based on a questionnaire collected through the 1ka.si platform. It was distributed partly among students of the University of Economics as part of their classes, and partly through social networks (Facebook questionnaire groups, Discord servers, Instagram, WhatsApp) and e-mail.

A preliminary phase (4–8 November 2024, 4 respondents) tested the clarity of the questions; a second phase (19–22 November 2024, 39 respondents) was run as a classroom exercise. As respondents found the questionnaire too long, several less important questions were removed so that completion took at most 20 minutes.

The entire questionnaire contains 30 questions after editing. In addition to basic demographic data and information about the use of social networks, both in the form of active use itself, and how often they use the social networks as a source of news and news. The following part of the questionnaire specifically addressed fake news and disinformation and verification of information coming from social networks. Respondents answered what tools they knew for verifying information and whether they verified information. If they answered that they used some tools, they were offered the opportunity to directly choose from the tools TinEye, InVID, Google Lens, RevEye, FotoForestics, Youtube DataViewer, Google Fact Check Tools, NewsGuard, Verifee, Demagog.cz, Manipulatori.cz, Snopes.com, Politifact.com.

The next part verified various false or true news and examined the respondent's reactions to them. The conclusion of the questionnaire focused on cognitive reflection, thinking and personality traits, where at the end there was an optional part from which respondents directly received answers with their personality profile.

All questions contained clear answers, instead of one question no texts were written anywhere. In addition to yes-no answers or frequency distinctions, a Likert scale was also used, ranging from 1 – I completely agree to 5 – I completely disagree.

2.1. False message detection

In one of the last parts of the questionnaire, we included several strange news stories, some of which were true and some false.

The false news stories included:

- Kamala Harris called Donald Trump Hitler (politifact.com)
- Floods in the Czech Republic were used by the government to intimidate citizens. (Facebook profile: Jan Skalický)
- Andrej Babiš pleaded guilty in the Bird's Nest case. (Invented placebo report)
- According to the new EU law, every family must accept a refugee. (Invented placebo report created based on novinky.cz)
- The Ukrainian army had Nazi symbols on their vehicles during the invasion in Kursk. (facebook.com)

True news included:

- Fiala says Czech Republic will continue to support Ukraine. (echo24.cz)
- Robert Fico appeared on Russian television, where he said he would accept Putin's invitation to celebrate Victory Day. (ČT24)
- Iran attacked Israel. (ČT24)

We asked users whether they had come across the news and if they had, they could indicate that they were not sure of the truth or directly indicate whether the news was true or not. The items were chosen to reflect news that was topical in autumn 2024, to combine domestic and foreign topics, and to include both items for which a fact-check was available and invented placebo reports, so that respondents could not rely on prior familiarity alone.

3. Methods

The data was imported from the 1ka.si server and then processed using Excel. Various functions and pivot tables were used for processing. Accuracy was defined as the number of correctly identified items divided by the number of items a respondent judged to be true or false, excluding “not sure” answers. Differences between tool users and non-users were tested using Welch’s t-test and the Mann–Whitney U test for accuracy rates, and a chi-square test for item-level correct/incorrect counts, with a significance level of 0.05.

4. Results

4.1. Descriptive statistics

Of 841 accesses to the questionnaire, 404 respondents answered at least one question and 290 completed all mandatory parts; we use these 290 respondents, as the final part was optional. In total, 143 men and 147 women completed the questionnaire. As shown in Figure 1, most respondents were aged 19–20 (88 and 86 respectively), with numbers decreasing for older groups (combined in ten-year bands in the graph, with a slight local maximum at 51–52); the oldest respondent was 60.

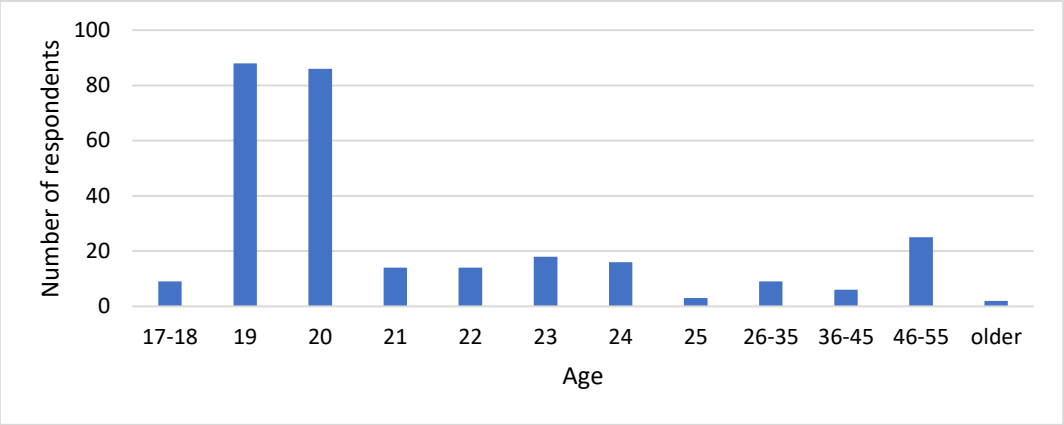


Figure 1. Age distribution of respondents Source: authors

In the questionnaire, we also asked about the municipality in which the respondents live, most (103) live in Prague, 121 in cities with a population of 1,000 and 66 in villages with a population of up to 1,000.

For the research itself, it is crucial how users use social networks and how they use social networks as a source of news. Table 1 summarizes the number of respondents who are on the respective social network at least once a day or whether they use the given network at least once a day as a source of news. It can be seen that the most popular social networks are Instagram and Youtube, both for active use in general and as a source of news.

Table 1. Using social networks at least once a day

Social network	Presence	As a news source
Instagram	248	212
Youtube	225	128
WhatsApp	175	47
Facebook	131	83
TikTok	106	80
Snapchat	84	16
Discord	66	14
X	50	47
Telegram	35	19
Pinterest	35	9
Reddit	19	11
LinkedIn	17	14
Zalo.me	1	2

Source: authors

4.2. Fact-checking webs use

The main part of the work is devoted to the use of fact-checking tools that can help the user determine whether a message is true or false. One of the questions in the questionnaire was What types of online information verification tools do you use? A summary of the answers is in Table 2. Of the 290 respondents, 156 do not use any online tools for verifying information, the remaining 134 use some of the tools.

Table 2. What types of online information verification tools do you use?

Tool type	Do not use	Use
Fact-checking webs	222	68
Browser extensions	219	71
Image verification tools	252	38
Video verification tools	269	21
None	156	134

Source: authors

We asked respondents who use at least one of the fact-checking methods about specific fact-checking applications – whether they know them and whether they use them. Among the most used is GoogleLens, which is used by 68 respondents and another 39 know about its existence. Applications that at least a quarter of respondents knew are summarized in Table 3.

Table 3. Fact-checking tools use

Tool	YES, I know and use	I know, but not use	Don´t know
Google Lens	68	39	24
Youtube DataViewer	18	39	71
Google Fact Check	27	26	75
Demagog.cz	34	19	76
Manipulatori.cz	24	22	82
Verifee	9	29	89

Source: authors

4.3. Detection of fake news based on the fact-checking tools used

For the data analysis in this section, we use data from users who determined whether at least one news item was true or false. There were 220 such users. At the same time, we look at whether the user is used to using some of the fact-checking tools. Of the 220 respondents in this section, 107 use fact-checking tools and 113 do not use these tools. These are therefore almost identical categories that will allow a good comparison of both categories. Table 4 summarizes how many news items users guessed correctly, divided according to whether they use fact-checking tools or not. Table 5, on the other hand, shows how many news items users guessed incorrectly, again divided according to the use of the tools.

Table 4. Correctly detected truthfulness of news items

Number of correctly detected news items	User uses tools	User do not use tools	Total
0	2	2	4
1	16	25	41
2	24	31	55
3	23	21	44
4	20	12	32
5	12	15	27
6	3	7	10
7	5	0	5
8	2	0	2

Source: authors

Table 4 shows that all users who were able to correctly guess all or most of the news are from among those who are commonly used to using fact-checking tools, however, their number is too small for a statistical conclusion. In the other parts, the numbers are balanced, although respondents who are used to the tools tend to win.

Table 5. False detected truthfulness of news items

Number of incorrectly detected news items	User uses tools	User do not use tools	Total
0	75	86	161
1	20	21	41
2	6	2	8
3	3	4	7
4	2	0	2
6	1	0	1

Source: authors

To compare the two groups, we calculated an accuracy rate for each respondent as the number of correctly identified items divided by the number of items the respondent judged to be true or false (excluding "not sure" responses). Mean accuracy was 88.2 % (SD = 22.6) for tool users and 90.7 % (SD = 20.2) for non-users. The difference was not statistically significant (Welch t-test, $p = 0.38$; Mann–Whitney U, $p = 0.32$), and a chi-square test on item-level correct/incorrect counts gave the same result ($\chi^2(1) = 1.94$, $p = 0.16$). Tool users tended to evaluate more news items, which inflates their absolute number of incorrect answers in Table 5 relative to their accuracy rate. We therefore find no evidence that using fact-checking tools is associated with better identification of true and false news.

4.4. Detection accuracy by individual news item

Accuracy varied substantially across the eight items (Table 6). Two false items — the alleged EU law requiring every family to accept a refugee (99 % correct) and the claim that floods were exploited by the government (96 %) — were near ceiling and therefore had limited discriminatory value.

Table 6. Identification success by individual news

Report	Truthfulness	Seen	Uncertain	Unambiguous	Correct	% correct (of unambiguous)	% uncertain (of seen)
Harris: Trump=Hitler	No	85	31	54	39	72.2 %	36.5 %
Fico on Russian TV	Yes	126	54	72	55	76.4 %	42.9 %
Iran attacked Israel	Yes	157	70	87	63	72.4 %	44.6 %
Floods abused by government	No	126	12	114	109	95.6 %	9.5 %
Babiš: Stork's nest	No	115	38	77	67	87.0 %	33.0 %
EU law: refugee	No	120	17	103	102	99.0 %	14.2 %
Ukrainian Nazi symbols	No	105	23	82	70	85.4 %	21.9 %
Fiala: will support Ukraine	Yes	189	29	160	152	95.0 %	15.3 %

Source: authors

The most difficult items were true news reports concerning foreign conflicts: "Iran attacked Israel" (72 % correct, 45 % unsure) and Fico's appearance on Russian television (76 % correct, 43 % unsure). Across all items, respondents expressed more uncertainty about true news (mean 34 % unsure) than false news (mean 23 %), and accuracy on true items (81 %) was lower than on false items (88 %), suggesting a tendency to reject rather than confirm news. This points to a limitation in item selection — future surveys should favour items of intermediate difficulty and balance the recognisability of true and false news.

5. Discussion and conclusion

The article summarizes the basic results of a survey regarding news verification, the use of fact-checking tools and their subsequent influence on the ability to detect false or true news.

We found no statistically significant difference in accuracy between users and non-users of fact-checking tools. Non-users were somewhat more likely to report being unsure about a news item, but this did not translate into a reliable difference in overall accuracy.

The results extend the findings of Sigmund and Korčák (2022), who identified a partial discrepancy between young respondents' confidence in their information literacy and their performance in practical tasks. The present study similarly suggests that the use of fact-checking tools should not be treated as a simple indicator of higher detection accuracy, since tool users were represented among both the most successful and the least successful respondents.

We are also aware that one item – concerning alleged Nazi symbols on Ukrainian army vehicles – touches on a politically sensitive and contested topic. Its results did not differ markedly from the other items, so it did not materially affect the overall comparison, but such items should be treated with particular care or excluded in future work.

A further limitation concerns the verification tools and sources themselves, which differ in their independence and in the technology on which they run. For news about international conflicts in particular, cross-checking against media from countries less directly involved in the given conflict – for example Indian outlets such as The Times of India or Hindustan Times alongside European sources – could make the assessment of truthfulness more robust, and the diversification of verification sources is a worthwhile direction for future work.

The main limitation of this study is its sample. Respondents were recruited mainly among university students and through social networks, resulting in a young, self-selected sample (most respondents aged 19–20) that is not representative of the general population. The selection of the tested news items was guided by their topicality in autumn 2024, a mix of domestic and foreign topics, and the availability of fact-checks, rather than by a representative sampling of disinformation themes. The findings should therefore be read as indicative rather than generalisable.

In future studies, a more in-depth statistical analysis – for example, principal component analysis of the observed responses – could help to identify latent patterns in how users evaluate news, beyond the descriptive and bivariate analyses presented here.

References

Google News Initiative. (n.d.). Google Fact Check Tools. https://newsinitiative.withgoogle.com/resources/trainings/google-fact-check-tools

Gregor, M., & Mlejnková, P. (2018). The best book about fake news, disinformation and manipulation (Nejlepší kniha o fake news, dezinformacích a manipulacích!!!). CPress. https://www.albatrosmedia.cz/tituly/40010211/nejlepsi-kniha-o-fake-news/

Gyüre, K. (2024). Media education in the context of infodemic and new media (Mediální výchova v kontextu infodemie a nových médií). Palacký University in v Olomouc. https://doi.org/10.5507/pdf.24.24464497

Kietzmann, J., Lee, L. W., McCarthy, I. P., & Kietzmann, T. C. (2020). Deepfakes: Trick or treat? Business Horizons, 63(2), 135–146. https://doi.org/10.1016/j.bushor.2019.11.006

NewsGuard Technologies, Inc. (n.d.). NewsGuard—Global Leader in Information Reliability. NewsGuard. https://www.newsguardtech.com

Sigmund, T., & Korčák, J. (2022). Fake news susceptibility of the young generation. In G. Chroust, P. Doucek, & V. Oškrdal (Eds.), *IDIMT-2022: Digitalization of Society, Business and Management in a Pandemic* (pp. 185–192). Trauner Verlag Universität. https://doi.org/10.35011/IDIMT-2022-185

Stahl, B. C. (2006). On the Difference or Equality of Information, Misinformation, and Disinformation: A Critical Research Perspective. Informing Science: The International Journal of an Emerging Transdiscipline, 9, 083–096. https://doi.org/10.28945/473

Tandoc, E. C., Lim, Z. W., & Ling, R. (2018). Defining "Fake News": A typology of scholarly definitions. Digital Journalism, 6(2), 137–153. https://doi.org/10.1080/21670811.2017.1360143

Verifee. (2024). Verifee – Your extension against disinformation. Verifee. https://www.verifee.ai/#1

BETWEEN TECHNOLOGICAL SIMPLIFICATION AND DIGITAL COLLABORATION: EVIDENCE ON SOLO AUTHORSHIP IN COMPUTER SCIENCE FROM THE DBLP DATABASE IN THE YEARS 2000–2024

Lubomír Štěpánek

Prague University of Economics and Business
Faculty of Informatics and Statistics
lubomir.stepanek@vse.cz

DOI: 10.35011/IDIMT-2026-237

Keywords

DBLP database; solo authorship; co-authorship patterns; scientific collaboration; digital collaboration; technological simplification; computer science; authorship trends; research team size; bibliometrics.

Abstract

Computer science research is increasingly embedded in a networked digital publication environment, where authors, venues, and research communities form expanding structures of collaboration. At the same time, technological simplification, digital research infrastructures, and, in recent years, AI-assisted workflows may reduce some barriers to individual research production. This creates a tension between two expectations: new technologies may support more independent work, while digital collaboration and research multicentricity may encourage larger author teams. This paper examines which tendency is more visible in the publication record of computer science. Using the DBLP database, the study analyzes 6,780,022 journal articles and conference proceedings published in the years 2000–2024. The analysis focuses on solo authorship, publication type differences, venue size, and authorship-size composition. The results show a strong long-term decline in solo-authored publications, accompanied by a steady increase in the mean number of authors per paper. Regression models confirm that later publication years are associated with lower odds of solo authorship, even after controlling for publication characteristics. Multinomial results further show that the change is not merely a shift from solo to minimally co-authored papers, but a broader movement toward medium-sized and larger teams. The findings suggest that computer science continues to move toward collaboration rather than a return to solo authorship.

1. Introduction

Scientific publications are not only records of research results, but also indicators of how knowledge production is organized. Authorship is one of the most visible traces of this organization: single-authored papers suggest individual research production, while multi-authored papers indicate cooperation, specialization, and coordination. Although co-authorship is not a perfect measure of

collaboration (Katz & Martin, 1997), it remains one of the most widely available indicators for studying large-scale changes in scientific work.

This issue is especially relevant in a networked digital publication environment, where scientific output can be understood as a system of authors, venues, institutions, and research communities. Co-authorship networks have long been used to study scientific collaboration (Newman, 2001), and large bibliographic datasets now make it possible to examine science as a complex, evolving system (Fortunato et al., 2018). Recent bibliometric work has also emphasized that expanding publication systems affect scientific visibility and the infrastructure of scholarly communication (Štěpánek, 2025), including peer-review capacity and decision processes (Štěpánek et al., 2024). Prior research shows a strong movement toward team-based knowledge production (Wuchty et al., 2007) and emphasizes the formation and evolution of research teams as a central feature of modern science (Milojević, 2014). At the same time, team size may matter for the type of contribution produced, since smaller teams are more often associated with disruptive work, while larger teams more often develop established research directions (Wu et al., 2019).

Computer science is a relevant field for examining this tension: specialization, complex methods, shared infrastructures, funding incentives, publication pressure, and conference-based publishing may encourage larger teams, while digital tools and AI-assisted workflows may reduce barriers for individual researchers (Khalifa & Albadawy, 2024). However, modern generative AI covers only the latest part of 2000–2024, so the earlier decline of solo authorship is more plausibly linked to broader collaborative and institutional changes.

This paper examines which tendency is more visible in the publication record of computer science. Using the DBLP database, a major bibliographic source for computer science (Ley, 2009), the analysis covers journal articles and conference proceedings published in the years 2000–2024. The year 2025 is excluded to avoid distortions caused by potentially incomplete indexing. The final dataset contains 6,780,022 records with at least one author. The main research question is: how has solo authorship in computer science changed in the years 2000–2024, and what does this reveal about the balance between technological simplification and digital collaboration?

The paper addresses this question through descriptive trends, publication-type comparisons, venue-size analysis, and authorship-size composition. Logistic regression models examine whether the decline of solo authorship remains visible after controlling for publication characteristics, while a multinomial model distinguishes between two-author papers, medium-sized teams, and larger author groups. The remainder of the paper describes the data and methods, presents the results, and discusses their implications for the changing organization of computer science research.

2. Research methodology

This section describes the data source, preprocessing workflow, statistical models, and software environment used to examine long-term changes in solo authorship in computer science.

2.1. Data source and acquisition

The analysis uses bibliographic metadata from the DBLP computer science bibliography. DBLP was selected because it is a field-specific source that provides structured information on computer science journals and conference proceedings (Ley, 2009). The complete DBLP dataset is distributed as an XML dump, which makes it suitable for reproducible large-scale bibliometric analysis. The original dblp.xml.gz file was processed in Google Colab using `Python`. This workflow was chosen because

it allowed the compressed XML file to be parsed directly without requiring a full local database installation (Kluyver et al., 2016).

2.2. Preprocessing and analytical sample

The extraction retained only DBLP records of type *article* and *inproceedings*, corresponding to journal articles and conference proceedings papers. Records were restricted to the years 2000–2024 and to publications with at least one listed author. The year 2025 was excluded in order to reduce the risk of distortions caused by incomplete indexing of the most recent publication year.

The final analytical dataset contains 6,780,022 publications. For each publication, a compact set of variables was retained. The analysis focuses on authorship size, publication type, time trends, and selected publication-level characteristics rather than textual content or citation links – the main variables used in the analysis are summarized in Table 1.

Table 1. Main variables used in the analysis

variable	description	use in analysis
`year`	publication year, 2000–2024	time trend
`year_c`	centered publication year, i.e. `year` – 2012	main time predictor
`type`	DBLP publication type: article or inproceedings	publication-type comparison
`n_authors`	number of listed authors	authorship size
`solo`	1 if `n_authors` = 1, otherwise 0	binary dependent variable
`team_size`	1, 2, 3–5, or 6+ authors	multinomial dependent variable
`pages_bin`	page-length category	control variable
`has_ee`	indicator for electronic edition metadata	control variable
`has_crossref`	indicator for cross-reference metadata	control variable
`log_venue_n`	logarithm of venue size in the dataset	venue-level control

The variable `solo` captures whether a publication is single-authored. The variable `year_c` denotes the centered publication year and was calculated by subtracting the midpoint of the observation period from the publication year, i.e. `year_c` = `year` − 2012. This centering improves interpretability of the regression intercept without changing the estimated time trend. For the multinomial analysis, the number of authors was additionally grouped into four categories: 1 author, 2 authors, 3–5 authors, and 6 or more authors. This grouping separates solo authorship from minimal co-authorship and from medium-sized and larger research teams.

2.3. Statistical modelling

The empirical analysis combines descriptive statistics and regression modelling. First, yearly trends in solo-authorship share, mean number of authors, and authorship-size composition are described. Second, binary logistic regression is used to model whether a publication is solo-authored, following standard approaches for binary outcomes (Hosmer et al., 2013). The main model for the i-th observation is specified as follows,

$$\begin{aligned}\operatorname{logit}(\mathrm{P}(\texttt{solo}_i = 1)) = {} & \beta_0 + \beta_1\,\texttt{year_c}_i + \beta_2\,\texttt{type}_i + \beta_3(\texttt{year_c}_i \times \texttt{type}_i) + \\ & + \beta_4\,\texttt{pages_bin}_i + \beta_5\,\texttt{has_ee}_i + \beta_6\,\texttt{has_crossref}_i + \\ & + \beta_7\,\log_{\texttt{venue_n},\,i} + \varepsilon_i\,. \end{aligned} \tag{1}$$

The coefficient β_1 for `year_c` captures the long-term change in the probability of solo authorship. The publication-type coefficient β_2 compares journal articles and conference proceedings, while the interaction between year and publication type, β_3, tests whether the time trend differs between the

two publication types. The remaining variables and their coefficients, $\beta_4, \beta_5, \ldots, \beta_7$ control for observable publication and venue characteristics.

Third, a multinomial logit model is used to examine whether the change is mainly a shift from solo to co-authorship in general or a movement toward larger author teams. The dependent variable distinguishes between four authorship-size categories, $\text{team_size}_i \in \{1, 2, 3\text{-}5, 6+\}$ authors. Solo-authored publications are used as the reference category. For each non-solo category k and the i-th observation, the model estimates

$$\log\left(\frac{\mathrm{P}(\texttt{team_size}_i = k)}{\mathrm{P}(\texttt{team_size}_i = 1)}\right) = \alpha_k + \gamma_k \, \texttt{year_c}_i + \delta_k \, \texttt{type}_i + \mathrm{X}_i \, \theta_k \,, \tag{2}$$

where k represents 2 authors, 3–5 authors, or 6+ authors, and X_i denotes the same set of control variables used in the binary model, while θ_k denotes the corresponding coefficients. The parameter α_k is the category-specific intercept, γ_k captures the effect of publication year for authorship category k relative to solo-authored publications, and δ_k captures the publication-type difference for category k relative to solo-authored publications. This specification follows the general framework of categorical data analysis for outcomes with more than two categories (Agresti, 2013). Data extraction and processing were carried out in `Python`, while descriptive summaries, logistic regression and multinomial logit modelling were performed in R (Venables & Ripley, 2002).

3. Results

The results show a clear long-term decline of solo authorship and a broader shift toward larger author teams in DBLP-covered computer science publications.

3.1. Descriptive development by publication type

Figure 1 presents the yearly share of solo-authored publications separately for journal articles and conference proceedings. The trend shows a strong decline in both publication types. At the beginning of the observation period, solo-authored publications still represented a substantial part of DBLP-covered computer science output. By 2024, however, solo authorship had become much less common. Conference proceedings are generally less likely to be solo-authored than journal articles, which is consistent with the project-based and collaborative role of conferences in computer science.

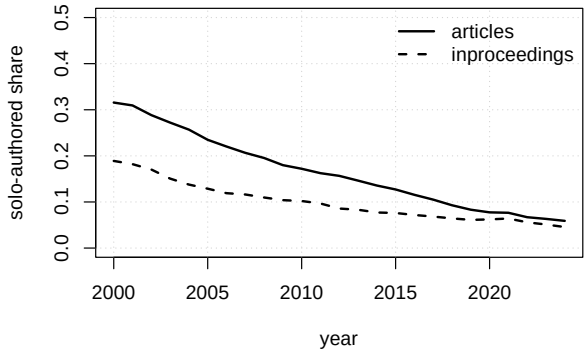


Figure 1. Solo-authored share by publication type, 2000–2024

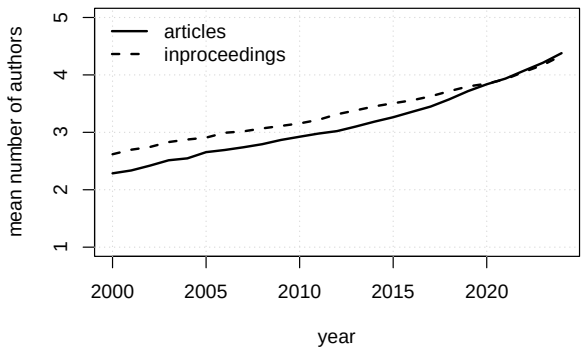


Figure 2. Mean number of authors by publication type, 2000–2024

Figure 2 shows the corresponding development in the mean number of authors per publication. The average number of authors increases steadily for both journal articles and conference proceedings. This confirms that the decline of solo authorship is accompanied by a broader increase in author-team

size. The pattern suggests a substantial change in the organization of computer science research: publications increasingly involve multiple contributors, and the typical author team has become larger.

3.2. Authorship-size composition

The authorship-size composition provides a more detailed view of the shift away from solo authorship. Figures 3 and 4 divide publications into four categories: one author, two authors, three to five authors, and six or more authors. Figure 3 shows the pattern for journal articles, while Figure 4 shows the pattern for conference proceedings.

Both figures show that the decline of solo authorship is accompanied by growth in larger authorship categories. The change is therefore not only a movement from one-author to two-author papers, but a broader shift toward medium-sized and larger teams. This pattern appears in both journal articles and conference proceedings, suggesting a general transformation in the organization of computer science research.

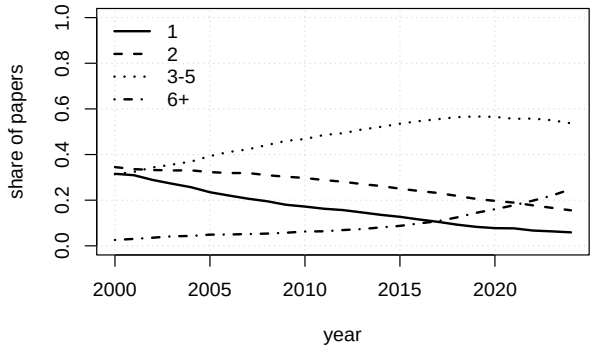


Figure 3. Authorship-size composition of journal articles, 2000–2024

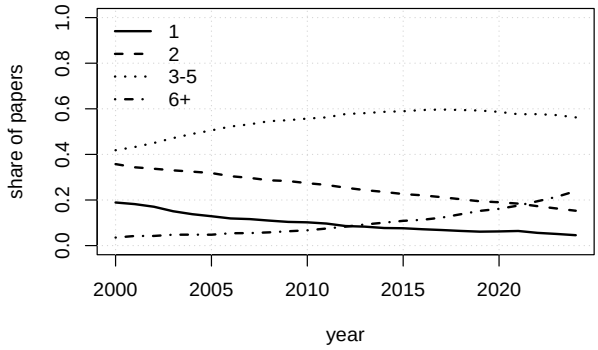


Figure 4. Authorship-size composition of conference proceedings, 2000–2024

3.3. Binary logistic regression results

The binary logistic regression confirms the descriptive findings. The dependent variable indicates whether a publication is solo-authored. As shown in Table 2, the coefficient for centered publication year is negative and statistically significant. In odds-ratio terms, each additional year is associated with about 8.2% lower odds of solo authorship (OR = 0.918, $p < .001$). This result remains visible after controlling for publication type, page length, electronic-edition metadata, crossref metadata, and venue size.

Table 2. Binary logistic regression predicting solo authorship

predictor	estimate	odds ratio (OR)	95% CI for OR	*p*-value
year_c	-0.0857	0.918	0.917-0.919	< .001
type: inproceedings	-0.5540	0.575	0.454-0.728	< .001
pages_bin 5-8	-0.6865	0.503	0.493-0.514	< .001
pages_bin 9-12	-0.9207	0.398	0.389-0.407	< .001
pages_bin 13-20	-0.9470	0.388	0.379-0.397	< .001
pages_bin 21+	-0.7764	0.460	0.446-0.475	< .001
has_ee	-0.2860	0.751	0.716-0.788	< .001
has_crossref	-0.3212	0.725	0.573-0.918	.008
log_venue_n	-0.2792	0.756	0.753-0.760	< .001
year_c vs. type: inproceedings	0.0303	1.031	1.028-1.033	< .001

The model also shows that conference proceedings are less likely to be solo-authored than journal articles, with about *42.5% lower odds* of solo authorship (OR = 0.575, $p < .001$). Larger venues are also associated with lower odds of solo authorship (OR = 0.756, $p < .001$), suggesting that more productive publication environments are more strongly connected to collaborative authorship. The positive interaction between year and inproceedings is statistically significant (OR = 1.031, $p < .001$), indicating that the decline occurs in both publication types, but with different slopes.

3.4. Grouped binomial model

As a robustness-style specification, a grouped binomial model was estimated on aggregated counts of solo and non-solo publications within combinations of year, publication type, page-length category, and metadata indicators. The results point in the same substantive direction as the individual-level logistic regression: later years are associated with lower odds of solo authorship (`year_c`: OR = 0.914, $p < .001$), and conference proceedings are less likely to be solo-authored than journal articles (`type`: inproceedings, OR = 0.516, $p < .001$). The positive interaction between year and publication type is also retained (OR = 1.037, $p < .001$), confirming that the decline occurs in both publication types but with different slopes.

Table 3. Grouped binomial model predicting solo-authorship share

predictor	estimate	odds ratio (OR)	95% CI for OR	*p*-value
`year_c`	-0.0904	0.914	0.913-0.914	< .001
`type`: inproceedings	-0.6608	0.516	0.466-0.572	< .001
`pages_bin` 5-8	-0.6735	0.510	0.506-0.514	< .001
`pages_bin` 9-12	-0.7788	0.459	0.455-0.463	< .001
`pages_bin` 13-20	-0.7116	0.491	0.487-0.495	< .001
`pages_bin` 21+	-0.4761	0.621	0.614-0.629	< .001
`has_ee`	-0.5702	0.565	0.555-0.576	< .001
`has_crossref`	0.0591	1.061	0.957-1.176	.259
`year_c vs. type`: inproceedings	0.0368	1.037	1.037-1.038	< .001

Note: Dependent variable: grouped counts of solo and non-solo publications. Reference publication type: article. The intercept is omitted from the table. Model AIC = 33,220.

3.5. Multinomial logit results

The multinomial logit model distinguishes between one-author, two-author, three-to-five-author, and six-or-more-author publications. Solo-authored publications serve as the reference category. Table 4 reports the key predictors for the non-solo categories.

Table 4. Selected multinomial logit results for authorship-size categories

outcome vs. solo	predictor	estimate	odds ratio (OR)	*p*-value
2 authors	`year_c`	0.0330	1.034	< .001
	`type`: inproceedings	0.4564	1.578	.026
3-5 authors	`year_c`	0.0827	1.086	< .001
	`type`: inproceedings	0.7183	2.051	< .001
6+ authors	`year_c`	0.1596	1.173	< .001
	`type`: inproceedings	0.8509	2.342	< .001

Note: Reference outcome: 1 author. The table reports selected key predictors. Page-length categories and metadata indicators were included as controls. Residual deviance = 915,586.9; AIC = 915,640.9.

The year effect increases across larger authorship groups. The odds ratio for year is 1.034 for two-author papers, 1.086 for three-to-five-author papers, and 1.173 for publications with six or more authors, each relative to solo-authored publications. This means that the long-term change is not simply a transition from solo to minimally co-authored papers. Instead, later years are increasingly associated with medium-sized and larger teams.

4. Discussion

The results show a clear long-term decline of solo authorship in DBLP-covered computer science publications. This decline appears in both journal articles and conference proceedings and remains visible after controlling for publication type, page length, metadata indicators, and venue size, suggesting a systematic feature of computer science publishing in the years 2000–2024.

The main interpretation is that collaboration-oriented forces dominate over the potentially individualizing effects of technological simplification. Importantly, the decline starts long before modern generative AI, so the pre-AI trend is more plausibly related to specialization, methodological complexity, shared infrastructures, funding structures, publication pressure, and multicentric collaboration. AI-assisted workflows may support individual scholars, but the publication record does not show a visible reversal of the long-term collaborative trend.

The authorship-size results strengthen this interpretation. The change is not merely a shift from one-author to two-author papers, but toward medium-sized and larger teams. This suggests a broader transformation in research organization, where computer science publications increasingly emerge from distributed groups, specialized roles, and collaborative infrastructures.

Differences between publication types and venues also matter. Conference proceedings are less likely to be solo-authored than journal articles, although the decline appears in both publication types. Similarly, the negative association between venue size and solo authorship suggests that larger venues may be connected to more established and collaborative research environments.

The findings should be interpreted cautiously. The analysis does not directly measure AI use, digital-platform use, funding structures, institutional collaboration, author contributions, or computer-science subfields. The decline may therefore partly reflect changing subfield composition, for example the growing weight of machine learning and data-intensive areas with larger teams, while theoretical areas may follow different authorship patterns. The 6+ category also does not distinguish ordinary larger teams from rare cases of hyper-authorship. Overall, the results suggest that technological simplification, closely related to AI tools, has not visibly reversed the long-term movement toward collaboration. In DBLP-covered computer science, the digital research environment is associated less with a revival of solo authorship than with the continued growth of team-based knowledge production.

5. Conclusions

This paper examined the development of solo authorship in computer science using DBLP journal articles and conference proceedings from the years 2000–2024. The study was motivated by a tension between technological simplification, which may support individual research work, and digital collaboration, specialization, and multicentricity, which may encourage larger author teams.

The results show a clear decline of solo authorship and a steady increase in the mean number of authors per publication. This pattern appears in both journal articles and conference proceedings and

remains visible in regression models after controlling for publication characteristics and venue size. The multinomial results further show that the change is not only a shift from solo to minimally co-authored papers, but a broader movement toward medium-sized and larger teams. The findings do not support causal claims about AI or specific digital tools, since these are not directly observed in the DBLP data. However, they show that the digital research environment has not coincided with a return to solo authorship. Instead, computer science publications increasingly reflect collaborative and team-based knowledge production. Future research could extend the analysis by adding subfields, institutions, countries, funding, citations, or direct indicators of AI and digital-tool use.

Acknowledgement and disclosure

This paper is supported by the grant IG410035 with no. F4/51/2025, which has been provided by the Internal Grant Agency of the Prague University of Economics and Business. AI-based language assistance, including ChatGPT 5.4, was used only for language polishing; all analyses, interpretation, and final responsibility remain with the author.

References

Agresti, A. (2013). Categorical data analysis (3rd ed.). John Wiley & Sons.

Fortunato, S., Bergstrom, C. T., Börner, K., Evans, J. A., Helbing, D., Milojević, S., Petersen, A. M., Radicchi, F., Sinatra, R., Uzzi, B., Vespignani, A., Waltman, L., Wang, D., & Barabási, A.-L. (2018). Science of science. Science, 359(6379), eaao0185. https://doi.org/10.1126/science.aao0185

Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). Applied logistic regression (3rd ed.). Wiley. https://doi.org/10.1002/9781118548387

Katz, J. S., & Martin, B. R. (1997). What is research collaboration? Research Policy, 26(1), 1–18. https://doi.org/10.1016/S0048-7333%2896%2900917-1

Khalifa, M., & Albadawy, M. (2024). Using artificial intelligence in academic writing and research: An essential productivity tool. Computer Methods and Programs in Biomedicine Update, 5, 100145. https://doi.org/10.1016/j.cmpbup.2024.100145

Kluyver, T., Ragan-Kelley, B., Pérez, F., Granger, B., Bussonnier, M., Frederic, J., Kelley, K., Hamrick, J., Grout, J., Corlay, S., Ivanov, P., Avila, D., Abdalla, S., & Willing, C. (2016). Jupyter notebooks — a publishing format for reproducible computational workflows. In F. Loizides & B. Schmidt (Eds.), *Positioning and Power in Academic* Publishing: Players, Agents and Agendas (pp. 87–90). IOS Press. https://doi.org/10.3233/978-1-61499-649-1-87

Ley, M. (2009). DBLP: Some lessons learned. Proceedings of the VLDB Endowment, 2(2), 1493–1500. https://doi.org/10.14778/1687553.1687577

Milojević, S. (2014). Principles of scientific research team formation and evolution. Proceedings of the National Academy of Sciences of the United States of America, 111(11), 3984–3989. https://doi.org/10.1073/pnas.1309723111

Newman, M. E. J. (2001). The structure of scientific collaboration networks. Proceedings of the National Academy of Sciences of the United States of America, 98(2), 404–409. https://doi.org/10.1073/pnas.98.2.404

Štěpánek, L. (2025). More is less? Exploring the dilution effect of growing publication network on citation rates in hard sciences. In IDIMT-2025: ICT in Business: AI Everywhere? Glory and Disgrace of AI. Trauner Verlag. https://doi.org/10.35011/IDIMT-2025-301

Štěpánek, L., Habarta, F., Malá, I., & Marek, L. (2024). Reviewers' fallacies: The optimal number of reviewers per manuscript and the surprising effect of "lazy" reviewers. In IDIMT-2024: Changes to ICT, Management, and Business Processes through AI. Trauner Verlag. https://doi.org/10.35011/IDIMT-2024-153

Venables, W. N., & Ripley, B. D. (2002). Modern applied statistics with S (4th ed.). Springer.

Wu, L., Wang, D., & Evans, J. A. (2019). Large teams develop and small teams disrupt science and technology. Nature, 566(7744), 378–382. https://doi.org/10.1038/s41586-019-0941-9

Wuchty, S., Jones, B. F., & Uzzi, B. (2007). The increasing dominance of teams in production of knowledge. Science, *316*(5827), 1036–1039. https://doi.org/10.1126/science.1136099

AI & DATA-DRIVEN SUPPLY CHAIN MANAGEMENT

GOVERNANCE OF SOCIAL VALUE AI FOR PUBLIC PROCUREMENT AND SUPPLY CHAINS

Radoslav Delina

Technical University of Kosice, Slovakia
Faculty of Economics
radoslav.delina@tuke.sk

Michaela Sulova

Catholic University in Ruzomberok, Slovakia
Faculty of Theology
michaela.sulova@ku.sk

Patrik Koba

Bratislava University of Economics and Business, Slovakia
Faculty of Business Economics with seat in Kosice
patrik.koba@euba.sk

DOI: 10.35011/IDIMT-2026-247

Keywords

Artificial Intelligence; public procurement; social value; social enterprises; supply chain governance; semantic matching.

Abstract

Public procurement is increasingly expected to support not only economic efficiency but also social inclusion, employment of disadvantaged groups and development of socially responsible supply chains. However, social value in procurement is still often treated as a formal requirement, based on supplier status, declarative criteria or weakly verified promises. This paper proposes a conceptual framework for governance of Social Value AI in public procurement and supply chains. The focus is deliberately narrowed to measurable and verifiable social value related to inclusion of disadvantaged groups and the participation of social enterprises or other suppliers able to prove comparable social contribution. Based on socially responsible public procurement, supplier selection and recent AI-in-procurement literature, the paper develops a lifecycle-oriented framework connecting need identification, market intelligence, tender design, multi-criteria evaluation, contract monitoring and public accountability. The contribution is a Social Value Intelligence architecture integrating procurement data, social enterprise registers, employment-related verification data, semantic matching, multi-criteria assessment and explainable decision support. The paper argues that AI should not replace procurement judgement but can strengthen transparency, comparability and

governance of social value if supported by auditable data infrastructure, human oversight and safeguards against bias and gaming.

1. Introduction

Public procurement belongs to the strongest instruments through which public authorities can influence the structure of markets and the behaviour of supply chains. In the traditional view, procurement is mainly a mechanism for buying goods, services or works under transparent and competitive conditions. In the broader strategic view, however, procurement is also a demand-side instrument for supporting innovation, sustainability, inclusion and better use of public resources. This is important especially in the context of supply chain management, where the buyer does not only select a supplier, but also shapes requirements, qualification conditions, performance indicators and the market opportunities for different types of suppliers.

The social dimension of public procurement is not new, but its practical implementation is still problematic. The European Commission guidance on socially responsible public procurement emphasises that public buyers can take social considerations into account in different phases of the purchasing process, if they are related to the subject matter of the contract and applied proportionately (European Commission, 2021). The UK Social Value Model gives a more operational example of how social value can be included in award processes and reporting (Cabinet Office, 2025). These approaches show that social value is becoming part of mainstream procurement policy. At the same time, the operational question remains open: how can the contracting authority know which social value is relevant, how can it find suppliers able to deliver it, and how can it verify that the declared value was really delivered.

This paper addresses this question from the perspective of artificial intelligence and data-driven supply chain governance. The paper is connected with earlier research on the participation of registered social enterprises in Slovak public procurement, where the descriptive evidence indicated that the participation of such enterprises is growing, but their success is concentrated in a limited number of firms and categories. The present paper does not repeat this descriptive question. Instead, it uses it as a motivation for a more general problem: how to move from formal participation of social enterprises towards measurable, verifiable and governable social value in public procurement supply chains.

The objective of the paper is to propose a conceptual framework for governance of Social Value AI in public procurement. The focus is intentionally narrower than the very broad concept of social value used in some policy documents. The paper concentrates on social value related to the employment and integration of disadvantaged groups, participation of social enterprises, and comparable evidence-based social contribution of other suppliers. This narrower scope is useful because it is closer to available administrative data, more suitable for verification, and more relevant for the development of AI-supported procurement services.

2. Literature survey

The literature relevant for this paper can be divided into four connected streams. The first stream concerns socially responsible public procurement and the use of procurement to create public value. Social procurement is usually understood as the inclusion of social objectives into procurement decisions and contract delivery. The European Commission (2021) stresses that social considerations may include employment opportunities, decent work, social inclusion, accessibility and ethical trade.

Selviaridis et al. (2023) show that strategic public procurement can create social value through supplier selection, market shaping and contract management, but that this requires active design of procurement processes and not only formal compliance. For this paper the important point is that social value is not only an attribute of the supplier; it is a result of the interaction between the public need, contract design, supplier capability and delivery monitoring.

The second stream concerns social enterprises and social impact supply chains. Social enterprises are usually expected to combine economic activity with social mission, but their position in supply chains differs from conventional suppliers. Zhang et al. (2024) show that social enterprises in supplier selection consider not only price, quality and delivery, but also value alignment, relationship commitment, resource complementarity and possibilities of social value creation. This is close to the logic of public procurement, where the buyer may need to consider whether the supplier can generate additional social effects. However, the same literature also shows that social value is difficult to compare and that suppliers may differ strongly in capacity, credibility and ability to report outcomes. Therefore, social value assessment cannot be reduced to a binary question whether the supplier has or has not a specific institutional status.

The third stream is the literature on supplier selection and multi-criteria decision making. Supplier selection has long been treated as a multi-criteria problem, because price alone cannot capture delivery capability, quality, risk, innovation potential or sustainability (Weber et al., 1991; Ho et al., 2010). With the development of sustainable and responsible supply chains, social and environmental criteria have been added to the supplier selection problem. In the context of social procurement, this creates a specific challenge. Social criteria can be used as qualification conditions, award criteria, contract clauses or monitoring indicators. Each option has different consequences. If social criteria are used only as a minimum requirement, they may open the market to eligible suppliers but do not necessarily compare higher and lower social contribution. If they are used as award criteria, they require more precise weighting, evidence and auditability.

The fourth stream concerns AI in procurement and public sector decision support. The OECD (2025) notes that AI can support public procurement by anticipating market trends, matching suppliers to needs, setting price benchmarks and supporting integrity controls. Recent work also discusses AI for sustainable public procurement and indicates opportunities in text analysis, supplier discovery, risk detection and contract monitoring (Andhov, 2025). At the technical level, semantic search and embedding models are relevant because procurement data are often textual, multilingual and weakly standardised. Current embedding benchmarks such as the Massive Text Embedding Benchmark compare large numbers of models across retrieval, classification, clustering and semantic similarity tasks (MTEB, 2026). Commercial embedding and language models can be used in hosted environments, while open models such as multilingual E5, BGE-M3 or Qwen3 Embedding can be deployed in more controlled infrastructures (Wang et al., 2024; Xiao et al., 2024; Zhang et al., 2025). Nevertheless, general benchmark performance is not sufficient for public procurement. Domain-specific validation is necessary because tender titles, CPV codes, legal clauses and social criteria have their own language and institutional meaning.

Together these streams support the main argument of the paper. Social value procurement needs more than legal permission and general policy guidance. It needs a data and AI-enabled governance infrastructure able to identify social value potential, discover suppliers, structure criteria, support evaluation and monitor delivery. AI is not proposed here as an automatic decision maker. It is proposed as an analytical and governance layer that can increase transparency, comparability and learning in a supply chain environment where social value is otherwise difficult to observe.

3. From social enterprise participation to AI-enabled social value governance

The starting point of the proposed framework is the distinction between formal participation and measurable social value. Registered social enterprises represent an important institutional category because their legal status is relatively identifiable and because they are directly connected with social economy policy. In the Slovak context, this is also important because previous data show that registered social enterprises are present in electronic public procurement, but that the distribution of their participation and success is highly concentrated. This concentration does not automatically mean a problem. Some enterprises may simply be more capable, better organised or more experienced in tendering. However, it shows that the policy objective of social procurement cannot be evaluated only by counting contracts awarded to this category of firms.

The broader governance problem is that social value can be generated also by other organisations. A commercial supplier may employ disadvantaged persons, cooperate with sheltered workshops, create local integration opportunities, or include a social enterprise in its supply chain. A non-profit organisation or a local consortium may also generate relevant social value. Therefore, the paper adopts a dual view. On one side, registered social enterprises remain an important and visible group of suppliers that should be supported. On the other side, social value should be assessed as a measurable and contract-related contribution, not only as an institutional label.

This view connects social procurement with supply chain management. If a public buyer wants to use procurement for social inclusion, it has to understand the supply market, the subcontracting possibilities, the available local and national capacities, and the risk that formal criteria will not create real impact. AI can support this task by transforming fragmented data into a more structured view of the market. The contribution of AI is not that it defines social value alone. The contribution is that it can help to identify where social value is possible, which suppliers may deliver it, which criteria are comparable, and where verification is weak.

Table 1. Shift from formal participation to AI-enabled social value governance

Logic	Traditional social procurement problem	AI-enabled governance interpretation
Supplier status	The supplier is or is not a registered social enterprise.	Supplier status is one verifiable signal, but it is combined with evidence about delivered social contribution.
Market visibility	Public buyers know only visible or historically successful suppliers.	Semantic and data-driven market intelligence identifies potential suppliers and social value capacities across the supply chain.
Criteria	Social criteria are often declarative or copied from templates.	Criteria are selected according to contract relevance, measurability, verification and expected additionality.
Evaluation	Social value is either a minimum condition or a weakly weighted narrative promise.	Multi-criteria assessment compares price, quality, risk and verifiable social contribution.
Monitoring	Social value is checked only formally or not at all after award.	Contract delivery is monitored through indicators, evidence and anomaly signals.

Source: authors

Therefore, social value in this paper is defined as a measurable, verifiable and contract-relevant contribution to the inclusion of disadvantaged groups, especially through employment, work integration, accessibility of services or participation of social economy actors in supply chains. This definition is narrower than general ESG or public value concepts, but it is more suitable for

procurement operations. It can be linked with administrative data, supplier profiles, contract clauses and monitoring indicators.

4. Social value criteria: from declarative requirements to comparable assessment

Social value criteria have to be translated into a structure which is understandable for procurers, suppliers and reviewers. In many procurement situations, the contracting authority knows that it wants to support social inclusion, but it does not know how to express this requirement without creating legal uncertainty, excessive administrative burden or formalistic promises. This is the reason why criteria should be treated as a multi-dimensional assessment problem and not as a single statement.

Four types of criteria can be distinguished. Input-based criteria describe the resources or capacities that the supplier already has, such as the share of disadvantaged workers, the status of a social enterprise, certified work integration capacity or cooperation with a social economy actor. Process-based criteria describe how the supplier will deliver the contract, for example by using inclusive recruitment, mentoring, training or adapted working conditions. Output-based criteria describe direct deliverables, such as number of working hours provided by disadvantaged workers, number of training places or number of supported employees. Outcome-based criteria describe the more difficult but more important effects, such as longer-term employment integration, improved employability or better access to services.

For practical procurement purposes, not all criteria can be used in the same way. Some criteria are suitable as qualification or eligibility conditions. Others are better used as award criteria, because they allow comparison among suppliers. Some criteria should be transformed into contract performance clauses and verified during implementation. This distinction is important because social value loses credibility when a criterion is used in the wrong phase of procurement. A very ambitious outcome criterion may sound attractive in the tender, but if it is impossible to verify, it will not improve governance.

The paper therefore proposes a multi-criteria social value assessment framework. The purpose is not to create a universal monetary value of social impact. Such monetary conversion may be useful in some policy contexts, but it can also create false precision. A more useful approach for procurement is to assess social value along several dimensions: relevance, intensity, credibility, sustainability, additionality and risk. Relevance captures whether the proposed social value is connected with the contract. Intensity captures the scale of the contribution. Credibility captures whether the claim is supported by reliable data or verification. Sustainability captures whether the contribution is temporary or has a longer-term integration effect. Additionality captures whether the value would probably occur without the contract. Risk captures the probability of non-delivery, gaming or purely formal compliance.

The verification problem is central. Social value procurement cannot rely only on self-declaration. At the same time, it cannot require public buyers to process sensitive personal data about disability, unemployment or social status. The proposed solution is to create a verification layer based on aggregated indicators, certification and controlled data exchange. For example, the Social Insurance Agency or labour market institutions may hold information relevant for verifying employment of disadvantaged persons, but the procurement system should not expose personal records. It should only provide certified indicators, such as whether the supplier fulfils an agreed threshold or whether a claimed employment contribution is consistent with administrative evidence.

Table 2. Proposed multi-criteria social value assessment dimensions

Dimension	Meaning in procurement	Possible evidence or data source
Relevance	Connection between the proposed social value and the subject matter, delivery mode or local context of the contract.	Tender object, CPV, contract location, description of delivery process, previous similar contracts.
Intensity	Scale of the proposed contribution to inclusion of disadvantaged groups.	Number or share of disadvantaged workers, work hours, training places, subcontracting volume to social economy actors.
Credibility	Degree to which the social value claim can be objectively verified.	Register of social enterprises, aggregated employment records, certifications, previous delivery reports, third-party verification.
Sustainability	Expected duration and stability of the social effect.	Contract duration, employment continuity, retention indicators, repeated cooperation, integration plans.
Additionality	Extent to which the effect is additional to what would happen without the contract.	Baseline employment, historical supplier profile, counterfactual reasoning, comparison with ordinary delivery model.
Risk	Probability that the declared value will not be delivered or will be delivered only formally.	Past performance, inconsistency in declarations, sudden changes in status, abnormal price/social value combinations.

Source: authors

This logic is also relevant for AI. Models used for assessment should not learn directly from sensitive personal data if the same function can be achieved through aggregated and certified signals. The AI system should support the contracting authority by comparing evidence, detecting inconsistencies and explaining uncertainty, not by making hidden judgements about vulnerable persons.

5. AI-supported social procurement across the procurement lifecycle

5.1. Need identification and problem definition

The strongest place for social value is often before the tender is announced. If social value is considered only at the evaluation stage, it is usually reduced to a weighting exercise. In the problem definition phase, AI can support the contracting authority in identifying whether the procurement need has social value potential. Text classification models can analyse the object of procurement, CPV category, contract location and previous similar contracts and indicate whether the need is suitable for work integration, local social enterprise participation, accessibility improvements or the application of social clauses.

This task can be approached as a supervised or weakly supervised classification problem. Historical contracts where social criteria were applied can serve as an initial training set. However, given the relatively limited scale of labelled data in the Slovak environment, the model should combine available labels with rule-based knowledge and expert feedback. In addition, semantic similarity approaches based on large language models can be used to identify structurally similar procurement cases across datasets.

In empirical application, such an approach can be tested on national procurement data by comparing procurement objects, CPV categories and supplier profiles. The objective is not to translate individual tender titles into illustrative examples, but to evaluate whether semantic similarity and classification

models can systematically identify procurement areas where social value criteria or social enterprise participation could be considered.

5.2. Market intelligence and semantic supplier discovery

The second use case is market intelligence and supplier discovery. Traditional discovery through CPV codes, supplier lists or previous awards is not sufficient because social value suppliers may be invisible, regionally dispersed, weakly described or active under categories that do not fully describe their real capabilities. CPV codes are important for standardisation, but they are not enough for semantic understanding. A social enterprise that provides cleaning, maintenance, simple construction support or IT refurbishment may be discoverable under different codes and descriptions.

AI can therefore combine CPV-based filtering with semantic search. Embedding models transform tender titles, supplier descriptions, catalogues and previous contract objects into vectors that can be compared by similarity. In practice, this means that a contracting authority can enter a procurement need and receive not only suppliers with the same CPV code, but also suppliers whose previous activities are semantically similar. This is especially important for social enterprises because their official descriptions may not follow the same language as procurement classifications.

Commercial models such as OpenAI text embedding models, Google Gemini embeddings, Cohere Embed or Voyage embeddings may be useful where cloud services are acceptable. Open-source alternatives are important for public sector environments where cost, data sovereignty, auditability or local deployment matter. Multilingual E5 and BGE-M3 are especially relevant because they support multilingual retrieval and can be deployed in controlled infrastructure. BGE-M3 is also interesting for procurement use cases because it combines dense, sparse and multi-vector retrieval and supports long inputs, which can be useful when comparing short tender titles with longer supplier descriptions or contract documents (Chen et al., 2024). The MTEB benchmark is a useful starting point for comparing embedding models across retrieval, classification, clustering and semantic similarity tasks, but it should not be treated as a final procurement benchmark (Muennighoff et al., 2023).

A Slovak public procurement benchmark should test semantic retrieval on Slovak and Czech tender titles, CPV categories, supplier descriptions and legal terminology.

5.3. Tender design and social clause generation

The third use case concerns tender design. AI can search previous procurement documents, identify social clauses used in comparable contracts and propose alternative formulations. This can help smaller municipalities and less experienced contracting authorities because the legal and administrative risk of using social criteria is often a barrier. A retrieval-augmented generation approach is suitable here. The model should not generate clauses from general language knowledge only. It should retrieve approved templates, legal guidance, previous tenders and evaluation practices, and then generate a draft with references to the retrieved sources.

Such a system could also warn the buyer when a proposed criterion is probably too vague, not connected with the subject matter, difficult to verify or discriminatory. This is one of the most practical roles of AI in social procurement. It can reduce the cognitive and legal cost of using social criteria, but at the same time it can increase standardisation and auditability.

5.4. Evaluation and multi-criteria decision support

The fourth use case is evaluation. AI should not replace the procurement committee. Its role is to structure the comparison between price, quality, risk and social value. If a supplier promises

employment of disadvantaged persons, the system should show how this promise is measured, which evidence supports it, whether it is relevant to the contract and how it compares with other offers. Multi-criteria decision support methods, such as weighted scoring, analytic hierarchy logic or outranking approaches, can be combined with AI-based extraction and verification of criteria.

An important issue is the trade-off between price and social value. If social value is very weakly weighted, it becomes symbolic. If it is weighted without evidence, it may distort competition. The proposed framework therefore treats social value assessment as an explainable process. The system should not only provide a final score. It should provide a decomposition of the score, indicate missing evidence and show sensitivity to weights. This is necessary for review, appeal and trust.

5.5. Contract monitoring and learning

The fifth use case is post-award monitoring. Many social value models fail because the promised value is evaluated at tender stage but not verified during execution. AI can support monitoring by comparing promised indicators with contract reports, invoices, employment-related aggregated signals, site reports or other evidence. Anomaly detection can identify situations where the declared social value is inconsistent with contract size, delivery timing, employment indicators or previous supplier behaviour.

Monitoring should also create a learning loop. If a certain social clause repeatedly fails, the system should recommend a different formulation. If a certain type of contract is suitable for social value but receives no bids, the system should recommend market consultation or supplier capacity building. If the same small group of suppliers dominates all social value contracts, the system should signal a possible market concentration problem. In this sense, AI supports not only individual procurement decisions but the governance of the whole social procurement supply chain.

Table 3. AI support across the social procurement lifecycle

Procurement phase	AI-supported function	Relation to social value
Problem definition	Classification of procurement needs and detection of social value potential.	Shows where inclusion-oriented criteria can be relevant already before tender design.
Market intelligence	Semantic supplier discovery using CPV, text embeddings and supplier profiles.	Finds social enterprises and other suppliers with similar capabilities and verifiable social contribution.
Tender design	Retrieval of comparable clauses and generation of contract-specific social criteria.	Reduces legal uncertainty and supports criteria that are relevant and verifiable.
Evaluation	Multi-criteria decision support and explainable decomposition of social value evidence.	Allows comparison of price, quality, risk and social contribution without black-box scoring.
Monitoring	Verification, anomaly detection and learning from delivery outcomes.	Turns promised social value into auditable performance and improves future procurement design.

Source: authors

6. Social Value Intelligence architecture

The central framework of the paper is the Social Value Intelligence architecture. The term intelligence is used here not as a synonym for automation, but as a combination of data, analytical models, human

judgement and institutional governance. The architecture consists of four layers: data layer, AI analytical layer, decision and governance layer, and interaction layer. The data layer has to go beyond procurement records. EKS and national procurement data provide information about contracts, CPV codes, estimated values, bids, winners and prices. The register of social enterprises provides institutional identification. Company registers, financial statements and business classifications provide information about firm size, economic stability and sector. Employment-related administrative data could provide aggregated verification of social contribution. Socio-economic data should not be limited only to regional context, because suppliers and employment relations may operate nationally. However, local indicators remain important for understanding where social value is more needed.

The AI analytical layer integrates several types of models. Classification models detect social value potential in procurement needs. Embedding models support semantic matching between contracts and suppliers. Information extraction models identify social clauses, declared commitments and measurable indicators in tender documents. Recommendation models suggest suppliers, clauses or market consultation strategies. Anomaly detection models identify unusual patterns in social value claims, supplier concentration, pricing or delivery evidence. Large language models can support explanation, drafting and interaction, but should be connected with retrieval and controlled knowledge bases.

The decision and governance layer transforms analytical outputs into procurement-relevant recommendations. It should show not only the recommended supplier or criterion but also the reason for the recommendation, confidence level, missing evidence and possible risks. The interaction layer connects the system with contracting authorities, suppliers and citizens. Procurers need support for defining needs and criteria. Suppliers need alerts and explanations why a contract is relevant for them. Citizens and local stakeholders need understandable information about what social value was expected and what was actually delivered.

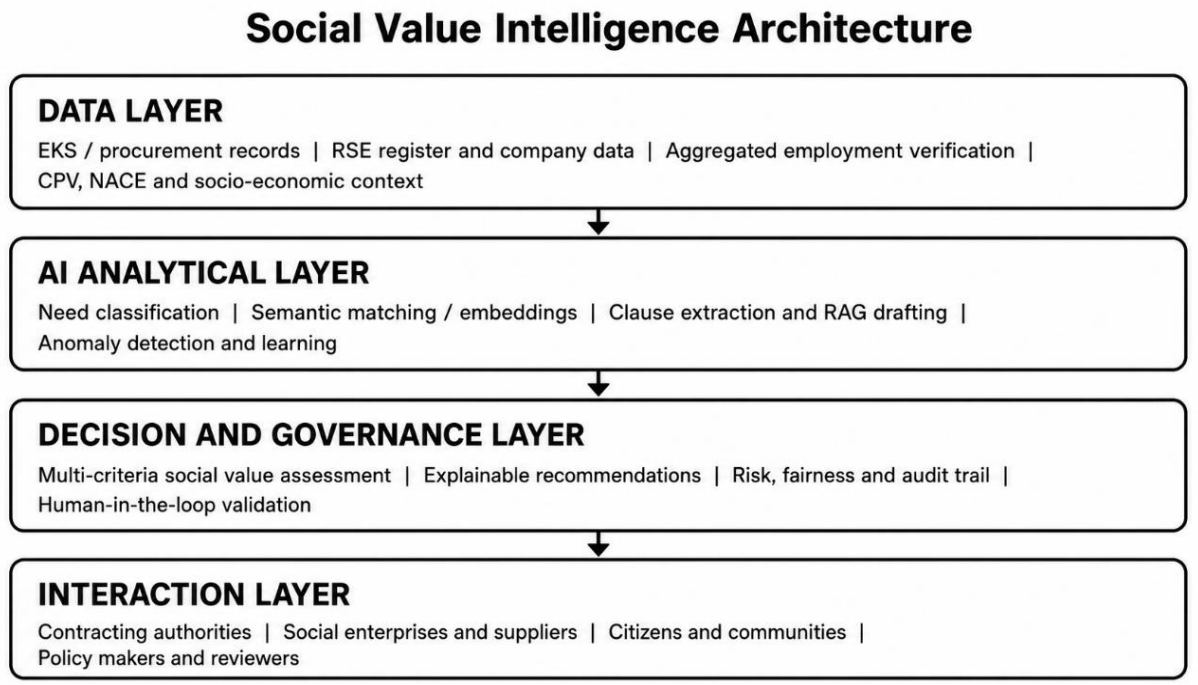


Figure 1. Proposed Social Value Intelligence architecture Source: authors

7. Participatory calibration of social value and public legitimacy

Social value is not a purely technical concept. It expresses what a community, public authority or society considers to be valuable beyond the immediate object of the contract. This means that a fully technocratic approach would be insufficient. At the same time, direct voting about suppliers or contract awards would create risks of populism, discrimination and legal uncertainty. The proposed solution is participatory calibration, not participatory award.

Participatory calibration means that citizens, community organisations, social services and local stakeholders can help to identify which types of social outcomes are important in a given context. For example, in a municipality with long-term unemployment, work integration may be prioritised. In a municipality with many persons with disabilities, accessibility and adapted services may be more important. In a neighbourhood regeneration project, local employment, community services and social enterprise participation may be relevant. These priorities can then inform the weighting or selection of criteria, but the final procurement decision remains subject to law, evidence and professional review.

AI can support this process by summarising citizen inputs, detecting common themes, clustering needs, simulating possible trade-offs and translating participatory priorities into procurement-relevant criteria. However, this requires safeguards. The model should not amplify majority preferences against vulnerable groups. It should also separate political preferences from legal procurement decisions. Participatory AI in social procurement is therefore not a tool for replacing expert judgement, but a tool for making social value priorities more explicit and transparent.

8. Risks, bias and safeguards in social value AI

The proposed framework has to be balanced by a discussion of risks. AI may increase social value, but it may also make existing problems stronger. If supplier recommendations are trained mainly on previous winners, the system may recommend the same dominant suppliers and reduce opportunities for emerging social enterprises. If the model uses weak textual signals, suppliers may learn to write social value promises that look convincing but are not connected with real delivery. If sensitive data are used without proper governance, the system may create privacy and discrimination risks.

Another risk is the formalisation of social value. A supplier may fulfil the visible requirement but not the real purpose of the policy. For example, employment of disadvantaged persons may be declared, but the quality, stability and integration effect may be low. A firm may also change its formal status in order to gain access to reserved or socially oriented procurement without building a deeper social mission. These cases are difficult to judge only from documents. This is why the architecture combines status verification, contract relevance, performance monitoring and anomaly signals.

The safeguards should therefore include human-in-the-loop decision making, audit trails, explainable recommendations, fairness constraints, periodic validation of model outputs and clear separation between advisory AI and legally binding decisions. Public procurement is a high-trust environment. Even when AI is used only as decision support, the buyer should be able to explain why a supplier was suggested, why a criterion was proposed and what evidence was used. This is consistent with the broader direction of AI governance in the public sector, where transparency, accountability and risk management are becoming essential conditions for adoption.

The most important research challenge is the development of domain-specific validation. General LLM and embedding benchmarks are not sufficient. A Social Value AI system should be evaluated on tasks such as correct identification of social value potential, retrieval of relevant suppliers,

detection of vague or unverifiable criteria, consistency of scoring across similar tenders and ability to explain recommendations. Only such domain benchmarks can show whether AI improves procurement governance rather than only producing fluent text.

9. Conclusion

The paper proposed a framework for governance of Social Value AI in public procurement and supply chains. The central argument is that social procurement should not remain only at the level of formal participation of registered social enterprises or declarative social criteria. It should move towards measurable, verifiable and contract-relevant social value. Registered social enterprises remain an important part of this system, but other suppliers may also create social value if their contribution is evidenced and connected with the contract.

The proposed Social Value Intelligence architecture links procurement data, social enterprise and company registries, aggregated employment verification, semantic matching, multi-criteria assessment, explainable recommendations and post-award monitoring. AI is positioned as an analytical and governance layer rather than as an automatic decision maker. Its role is to help contracting authorities identify social value potential, discover suppliers, design better criteria, evaluate evidence and learn from contract delivery.

The framework also shows that technological design alone is not sufficient. Social value is normative, context-dependent and sensitive. Therefore, the use of AI has to be supported by data governance, human oversight, participatory calibration, auditability and safeguards against bias and gaming. Further research should develop domain-specific benchmarks for social value procurement, test semantic matching on real procurement data and examine how different social criteria influence competition, supplier participation and delivered social outcomes.

References

Andhov, M. (2025). Leveraging AI for sustainable public procurement: Opportunities and challenges. Frontiers in Sustainability, 6. https://doi.org/10.3389/frsus.2025.1603214

Cabinet Office. (2025). PPN 002: Guide to using the Social Value Model. UK Government. https://www.gov.uk/government/publications/ppn-002-taking-account-of-social-value-in-the-award-of-contracts/ppn-002-guide-to-using-the-social-value-model-html

Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. ACM Computing Surveys, 41(3), 1-58. https://doi.org/10.1145/1541880.1541882

Chen, J., Xiao, S., Zhang, P., Luo, K., Lian, D., & Liu, Z. (2024). BGE M3-Embedding: Multi-lingual, multi-functionality, multi-granularity text embeddings through self-knowledge distillation. arXiv. https://arxiv.org/abs/2402.03216

Delina, R., et al. (2022). Digital Transformation of the Social Economy (in Slovak). National project of the Institute of Social Economy.

European Commission. (2021). Buying social: A guide to taking account of social considerations in public procurement (2nd ed.). Publications Office of the European Union. https://op.europa.eu/en/publication-detail/-/publication/47c69b3a-cfcf-11eb-ac72-01aa75ed71a1

European Union. (2024). Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). Official Journal of the European Union.

Ho, W., Xu, X., & Dey, P. K. (2010). Multi-criteria decision making approaches for supplier evaluation and selection: A literature review. European Journal of Operational Research, 202(1), 16-24. https://doi.org/10.1016/j.ejor.2009.05.009

HM Government. (2012). Public Services (Social Value) Act 2012. https://www.legislation.gov.uk/ukpga/2012/3/contents

Muennighoff, N., Tazi, N., Magne, L., & Reimers, N. (2023). MTEB: Massive Text Embedding Benchmark. Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics, 2014–2037. https://doi.org/10.18653/v1/2023.eacl-main.148

OECD. (2025). Governing with Artificial Intelligence. OECD Publishing. https://www.oecd.org/en/publications/governing-with-artificial-intelligence_795de142-en.html

Selviaridis, K., Hughes, A., & Spring, M. (2023). How strategic public procurement creates social value: Evidence from UK anchor institutions. Public Management Review. https://doi.org/10.1080/14719037.2023.2277814

Wang, L., Yang, N., Huang, X., Yang, L., Rangan, M., & Wei, F. (2024). Multilingual E5 text embeddings: A technical report. arXiv. https://arxiv.org/abs/2402.05672

Weber, C. A., Current, J. R., & Benton, W. C. (1991). Vendor selection criteria and methods. European Journal of Operational Research, 50(1), 2-18. https://doi.org/10.1016/0377-2217(91)90033-R

Zhang, X., van Donk, D. P., Xiao, C., & Pullman, M. (2024). Finding the right one: Understanding the supplier selection process of social enterprises. International Journal of Operations & Production Management.

Zhang, Y., Li, M., Long, D., Xin, X., Lin, H., Yang, B., Xie, P., Yang, A., Liu, D., Lin, J., & Huang, F. (2025). Qwen3 embedding: Advancing text embedding and reranking through foundation models. arXiv. https://arxiv.org/abs/2506.05176

OPERATIONAL EFFICIENCY OPTIMIZATION OF SUPPLY CHAIN

Sharfraj Ahmed, Shalini Aggarwal, Prerna Dhawan

Apex Institute of Management, Chandigarh University
sarfrajahmed149@gmail.com; shaliniaggar@gmail.com;
Erprerna_0407@rediffmail.com

Kiran Sood

Chitkara Business School, Chitkara University, Punjab, India
Kiransood1982@gmail.com

Simon Grima

University of Malta
Faculty of Economics, Management and Accountancy
simon.grima@um.edu.mt

Erik Weiss

Bratislava University of Economics and Business, Slovakia
Faculty of Business Economics with seat in Kosice
erik.weiss@euba.sk

DOI: 10.35011/IDIMT-2026-259

Keywords

Supply Chain Management; operational efficiency; logistics; Artificial Intelligence; sustainability.

Abstract

The study aims to present a bibliometric review of literature on optimizing operational efficiency in supply chain management. It maps the intellectual structure, thematic evolution, and emerging research trends from 2005 to 2026, highlighting the growing importance of efficiency, optimization, and data-driven decision-making in modern supply chains. A systematic bibliometric methodology was applied to a dataset of 170 documents from 133 sources. Advanced techniques, including annual scientific production analysis, citation analysis, keyword co-occurrence, thematic mapping, and collaboration network analysis, were employed. Key indicators such as average citations per document, author productivity, country-wise contributions, and source impact were also examined. The findings indicate a strong growth trend, with an annual increase rate of 14.76%, reflecting rising academic and practical interest. Core research themes include supply chain management, logistics,

and operational efficiency. Emerging trends highlight the integration of artificial intelligence, machine learning, and sustainability within supply chain optimization practices.

1. Introduction

Optimization of supply chain in terms of operational efficiency is the process of ensuring the processes, use of resources, and decision process tools are optimized to ensure maximum output with minimum wastages (Bakioglu, 2025; Krynke, 2024). The systems of supply chains have experienced a remarkable change in the last few decades as the simple linear systems have changed into a complex dynamic networks comprising of various stakeholders including suppliers, manufacturers, distributors and customers. Globalization, technology and evolving consumer demands have contributed to this growing complexity (Gotelli et al., 2026). As a result, the older approach of supply chain management, which was grounded on the traditional supply chain models and reactive strategies, is no longer applicable. Rather, more integrative and adaptive optimization methods are sought that can work with uncertainties and optimize all the operations (Galkin et al., 2025; Basulo-Ribeiro et al., 2024).

The increased value of operational efficiency is further justified by the fact that research in the field continues to develop at a very high pace. Bibliometric analysis of 170 publications identified and published between 2005 and 2026 has shown a steady increase in scholarly publication, with an average increase of 14.76 every year (Hu et al., 2023). It is a sign that more researchers and practitioners are becoming interested in optimizing supply chain operations using sophisticated analytical and technological solutions. Another important indicator of a collaborative research environment, shown in the dataset, is the number of 564 authors and a high percentage of multi-authored articles, which creates an interdisciplinary and international impression of the field (Makhija et al., 2025).

Along with technological innovation, the issue of sustainability has turned out to be a major concern in the optimization of the supply chain. Increasingly organizations are forced to ensure that they achieve economic efficiency at the expense of the environment and society (Wang et al., 2024). This has seen the development of sustainable and green supply chain practices to minimize the amount of carbon released in the environment, waste reduced and the more efficient use of energy. Operational efficiency is now not only restricted to cost reduction but also to long-term sustainability and resilience (Bu, 2024; Bridgelall & Tolliver, 2024).

2. Research methodology

Bibliometric analysis is a quantitative research method used to examine the structure, patterns and development of scientific literature through the statistical analysis of publications, citations, authorship, keywords and collaboration networks (Karagoz, 2026). Unlike traditional narrative reviews, bibliometric analysis enables a more systematic and reproducible examination of large volumes of indexed research data. This approach is suitable for analysing an interdisciplinary field such as operational efficiency optimization in supply chain management, where research is distributed across logistics, operations management, data analytics, artificial intelligence and sustainability.

The Scopus database was used as the source of bibliographic data. The search covered the period from 2005 to 2026 and focused on publications related to supply chain management, operational efficiency and optimization. The search was conducted on [insert exact Scopus search date] using the

following search query: TITLE-ABS-KEY ("supply chain" AND "operational efficiency" AND (optimization OR optimisation OR optimize* OR optimism*))

Only English-language publications indexed in Scopus were included. The dataset included journal articles, conference papers, book chapters and review papers related to supply chain management, logistics, operational efficiency, optimization, artificial intelligence, data analytics and sustainability. Publications that were not directly related to supply chain management or did not address operational efficiency in a meaningful way were excluded.

The initial Scopus export contained 172 records. After data screening and cleaning, two records were removed because they did not meet the inclusion criteria or contained incomplete bibliographic information. The final dataset used for bibliometric analysis therefore consisted of 170 documents from 133 sources. The analysis was conducted using Biblioshiny, the web interface of the Bibliometrix R package. The study applied annual scientific production analysis, citation analysis, keyword co-occurrence analysis, thematic mapping and country-level scientific production analysis. In this regard, the current paper attempted to present systematic literature review of operational efficiency optimization in supply chain management. In order to lead the analysis, the study is organized in relation to the following research questions:

- RQ1: *Which countries contribute most to research on operational efficiency optimization in supply chain management, and what patterns of international scientific production can be identified?*
- RQ2: *What are the main themes and research clusters emerging from keyword co-occurrence and thematic mapping?*
- RQ3: *Which established and emerging research themes can be identified in the analysed literature on supply chain operational efficiency optimization?*
- RQ4: *What research gaps and future research directions can be identified for the further development of supply chain operational efficiency optimization?*

These research questions all deal with the structural, thematic, and evolutionary aspects of the discipline. The answers to them will help the study to gain a clear picture of the current research state and distribute practical information on future academic and practical evolution.

3. Data analysis

3.1. Main information

The final dataset consists of 170 publications covering the period from 2005 to 2026. The annual scientific production growth rate of 14.76% indicates a clear increase in academic interest in operational efficiency optimization within supply chain management. The average age of the documents is 2.39 years, which suggests that a considerable part of the analysed literature consists of recent publications. This confirms that the topic has gained stronger attention in the last few years, particularly in connection with digital transformation, artificial intelligence, data analytics and sustainability-oriented supply chain practices. The average number of citations per document is 12.82. However, this indicator should be interpreted with caution. Citation impact depends on the year of publication, the disciplinary context, the type of document and the time available for citations to accumulate. Therefore, the value should not be understood as a direct measure of moderate or strong academic impact, but rather as a descriptive indicator of the visibility of publications included in the dataset.

The dataset includes 564 authors, which indicates a broad and diverse research community. Only 20 documents were written by a single author, while the majority of publications were co-authored. The average number of co-authors per document is 3.43, and 28.24% of the documents include international co-authorship. These results suggest that research on operational efficiency optimization in supply chain management is collaborative, interdisciplinary and increasingly international in nature.

3.2. Word cloud

The word cloud presents the most frequently occurring terms in the analysed dataset. The most frequent term is “supply chain management”, with 49 occurrences, followed by “logistics” with 37 occurrences and “optimization” with 29 occurrences. Other frequently used terms include “supply chains” with 27 occurrences and “supply chain” with 20 occurrences. These keywords confirm that the analysed literature is strongly concentrated around the core concepts of supply chain management, logistics and process optimization. Performance-related terms are also visible in the dataset. The term “efficiency” appears 18 times, while “operational efficiency” appears 14 times. This indicates that the improvement of performance, resource utilization and operational processes represents a central concern of the analysed literature. At the same time, the occurrence of “sustainability” with 15 occurrences and “artificial intelligence” with 14 occurrences suggests that the field is gradually moving beyond traditional cost-efficiency approaches toward more technologically advanced and sustainability-oriented models of supply chain optimization. The presence of terms such as “decision making” further indicates the importance of analytical and strategic approaches in supply chain research. These results should not be interpreted only as a list of frequent keywords. Since terms such as “supply chain management” and “operational efficiency” are directly related to the search strategy, their frequency is expected. The more important finding is the connection between traditional supply chain concepts and emerging themes such as artificial intelligence, sustainability, data-driven decision-making and optimization. This suggests a shift from conventional logistics efficiency toward smarter, more adaptive and more sustainable supply chain systems.

3.3. Thematic map

Figure 1 presents the thematic structure of the analysed research field according to the degree of centrality and density of the identified themes. Centrality reflects the relevance of a theme within the overall conceptual network, while density indicates the level of internal development of a given theme. The thematic map shows that supply chain management, artificial intelligence and machine learning form an important basic theme. This indicates that AI-related technologies are increasingly connected with the core discussion on operational efficiency in supply chain management. These technologies support efficiency improvements mainly through better demand forecasting, inventory planning, logistics optimization, data-driven decision-making and process automation. Another important thematic area is represented by supply chains, efficiency and cost reduction. This cluster can be interpreted as a motor theme because it is both highly developed and highly relevant for the field. It reflects the traditional managerial objective of supply chain optimization, namely the reduction of costs and improvement of operational performance.

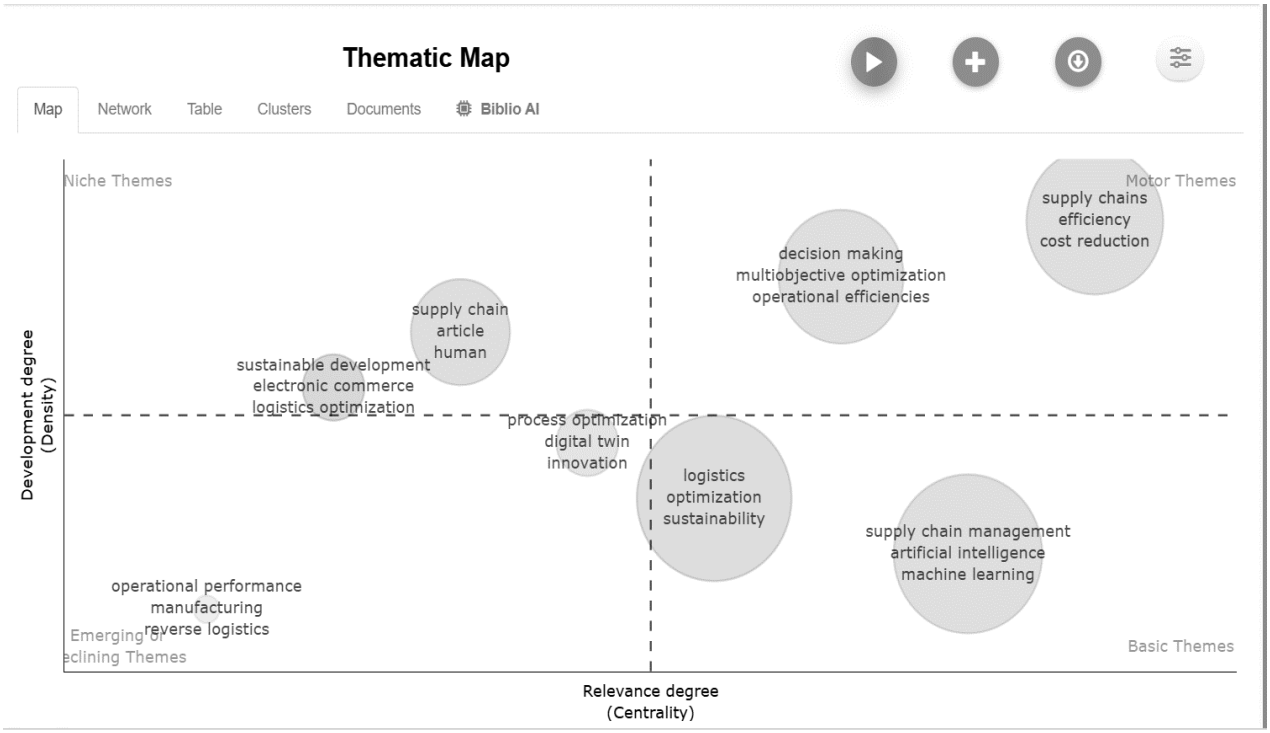


Figure 1. Thematic map Source: author's computation using Biblioshiny

The map also identifies themes related to decision making, multi-objective optimization and operational efficiencies. These topics show that the field is moving toward more complex optimization approaches, in which cost, time, quality, flexibility, risk and sustainability are considered simultaneously. Themes such as logistics, optimization and sustainability indicate the growing integration of environmental concerns into operational efficiency research. This suggests that contemporary supply chain optimization is no longer limited to cost reduction, but increasingly includes sustainable logistics, energy efficiency and carbon reduction.

Finally, digital transformation, predictive analytics, demand forecasting and digital twins appear as emerging themes. Although these themes are less mature than traditional supply chain management topics, they represent important future directions for research on smarter, more adaptive and data-drive supply chain systems.management topics, they represent important future directions for research on smarter, more adaptive and data-driven supply chain systems.

3.4. Countries scientific production

Figure 2 presents the geographical distribution of scientific production in the analysed dataset. China is the most active country, with 76 publications, followed by India with 49 publications and the United States with 29 publications. Among European countries, Portugal and the United Kingdom also show visible research activity, with 21 and 20 publications respectively. Italy, Malaysia, South Korea and Saudi Arabia contribute to the field to a lesser but still relevant extent. These results indicate that research on operational efficiency optimization in supply chain management is geographically concentrated in several highly active countries, but it also has an international character. The strong position of China and India may be associated with the importance of manufacturing, logistics and digital transformation in these economies. The presence of the United States and European countries confirms the relevance of the topic for advanced supply chain systems and technology-oriented research environments.

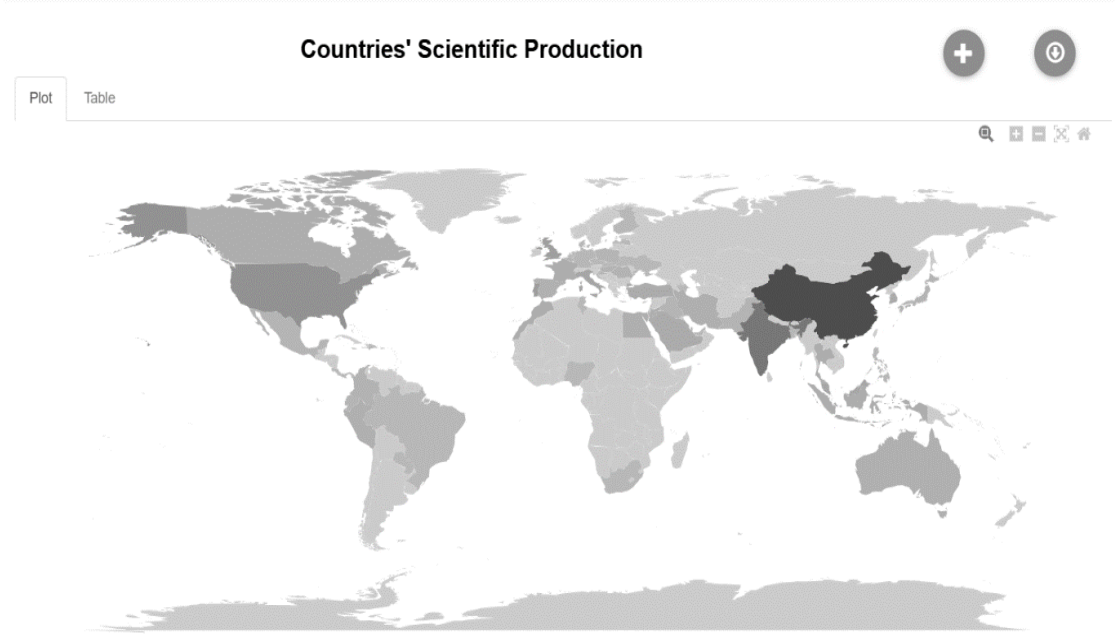


Figure 2. Country scientific production Source: author's computation using Biblioshiny

4. Implications of the study

The policymakers must focus on creating structures that allow the integration of new and innovative technologies like artificial intelligence, big data analytics, and digital platforms in the supply chain systems. The results show that these technologies are very important in enhancing the efficiency of operations, but they are not adopted uniformly regionally and across industries (Wen et al., 2025). The government policies and funding schemes should thus be concentrated into aiding technological adoption especially in the developing economies.

Another gap that the study mentions is in the connection between sustainability research and its practical application in the supply chains. Although the concept of sustainability has been identified as a significant research topic, there is a weak incorporation of sustainability into operational practices (Yeng et al., 2024). The policymakers ought to institute standard rules and controls that would motivate organisations to pursue sustainable operations as well as efficiency measures. This involves marketing of green logistics, transportation systems that are efficient in energy consumption and carbon mitigation measures (Wei et al., 2021).

Moreover, the fact that there are no standardized systems of classification and that metadata fields, including science categories, are incomplete, contributes to more general issues of organization of research and sharing knowledge (Younesse et al., 2025). Researchers ought to be encouraged to increase the standardization and accessibility of data in order to improve the quality and comparability of research findings, even by the policymakers and academic institutions (Sun et al., 2025).

Another key policy area is international collaboration. Current partnerships are limited in scope and depth. Expanding cross-border research through joint funding, academic exchanges, and international consortia can enhance knowledge sharing and drive innovation in supply chain optimization (Basulo-Ribeiro et al., 2024).

5. Conclusion

This paper presents a bibliometric analysis of research on operational efficiency in supply chain management, highlighting key themes and future research directions. Based on 170 documents published between 2005 and 2026, the study demonstrates the rapid growth and increasing relevance of this research field. The findings show that supply chain management, logistics and optimization represent the core thematic areas of the analysed literature, with operational efficiency serving as a key objective across different studies (Gotelli et al., 2026). The analysis also indicates a clear shift toward the adoption of advanced technologies, including artificial intelligence, machine learning and data analytics, which are transforming traditional supply chain systems into smarter, more flexible and data-driven networks. The results further reveal a high level of collaboration among researchers, indicating that the field is increasingly interdisciplinary and internationally connected. China, India and the United States are the most prominent countries in terms of scientific production, while leading journals play an important role in the dissemination of knowledge in this area (Hu et al., 2023). Despite these developments, several limitations and research gaps remain. These include the insufficient integration of theoretical approaches, limited differentiation of multi-objective optimization models and inadequate attention to real-time decision-making systems (Makhija et al., 2025).

These gaps provide opportunities for future research to strengthen the theoretical, methodological and practical relevance of the field. Overall, this study contributes to the existing literature by offering a systematic and structured overview of operational efficiency optimization in supply chains. It maps the current state of knowledge and provides useful insights for researchers, practitioners and policymakers (Bakioglu, 2025).

Acknowledgement

This research was funded by Slovak agency KEGA grant number 032EU-4/2024.

References

Bakioglu, G. (2025). Enhancing efficiency in railway freight logistics using a two-stage decision support technique with Q-rung orthopair fuzzy sets. Canadian Journal of Civil Engineering, 52(5), 770-795.

Basulo-Ribeiro, J., Pimentel, C., & Teixeira, L. (2024). Digital transformation in maritime ports: defining smart gates through process improvement in a Portuguese container terminal. Future Internet, 16(10), 350.

Bridgelall, R., & Tolliver, D. (2024). Transforming healthcare delivery with advanced air mobility: a rural study with GIS-Based optimization. Sustainability, 16(13), 5709.

Bu, L. (2024). Logistic resource allocation based on multi-agent supply chain scheduling using meta-heuristic optimization algorithms. Applied Artificial Intelligence, 38(1), 2362516.

Galkin, A., Samchuk, G., Kopytkov, D., & Thompson, R. G. (2025). Digital twins in logistics: a comprehensive bibliometric analysis for advancing smart cities and sustainable development. Discover Sustainability, 6(1), 853.

Gotelli, M., Ferraris, G. M., Ferrero, F., Giovannetti, A., & Bruzzone, A. G. (2026). Online food delivery and supply chain management by optimization of innovative models. International Journal of Food Engineering, 22(1), 27-35.

Hu, B., Guo, H., Tao, X., & Zhang, Y. (2023). Construction of digital twin system for cold chain logistics stereo warehouse. IEEE Access, 11, 73850-73862.

Karagoz, S. (2026). Resolving Conflicting Goals in Manufacturing Supply Chains: A Deterministic Multi-Objective Approach. Systems, 14(2), 126.

Krynke, M. (2024). Virtual simulation modeling as a key element of warehouse location optimization strategy. *Management* Systems in Production Engineering, (3 (32), 339-344.

Makhija, R., Aggarwal, S., & Singh, Y. (2025). AI-Driven Hybrid Learning for Sustainable Development: A Bibliometric Analysis and Systematic Literature Review. Prabandhan: Indian Journal of Management, 18(9). 29-53.

Sun, J., Li, X., & Chen, Z. (2025). An investigation into multi-objective decision-making in fresh cold chain supply chain networks within a dual distribution framework. Complex & Intelligent Systems, 11(9), 393.

Wang, J., Wang, L., & Han, H. (2024). A knowledge-driven cooperative coevolutionary algorithm for integrated distributed production and transportation scheduling problem. IEEE Transactions on Automation Science and Engineering, 22, 7435-7448.

Wei, Q., Sun, D., Gou, X., & Bai, C. (2021). Sharing economy for cost reduction and efficiency increase: the case of sharing E-commerce logistics. Complexity, 2021(1), 5573422.

Wen, H., AbouRizk, S., & Mohamed, Y. (2025). Gap analysis of digitalization levels in construction and manufacturing: a comparative study of construction 4.0 and industry 4.0. Journal of Construction Engineering and Management, 151(3), 04025001.

Yeng, F. F., Zainuddin, Z. M., & Pheng, H. S. (2024). Optimizing palm oil biomass supply chain logistics through multi-objective location-routing model. Malaysian Journal of Fundamental and Applied Sciences, 20(2), 247-265.

Younesse, O. U. A. H. B. I., & Soumia, Z. I. T. I. (2025). NW Logistics: System Architecture and Design for Sustainable Road Logistics. International Journal of Advanced Computer Science & Applications, 16(4).

GENERATIVE AI IN ESG COMMUNICATION AND SUSTAINABLE CORPORATE BRANDING: TRUST AND SYNTHETIC AUTHENTICITY

Martin Bachura, Magdaléna Freňáková, Tatiana Jurovčáková, Slávka Bocánová

Bratislava University of Economics and Business, Slovakia
Faculty of Business Economics with seat in Kosice
martin.bachura@euba.sk, magdalena.frenakova@euba.sk,
tatiana.jurovcakova@euba.sk, slavka.bocanova@euba.sk

DOI: 10.35011/IDIMT-2026-267

Keywords

ESG communication; generative AI; sustainable branding; stakeholder trust; synthetic authenticity; greenwashing.

Abstract

Generative artificial intelligence (AI) is reshaping how corporations produce and distribute ESG communication, yet existing frameworks in marketing and sustainability management have not kept pace with this shift. The aim of this paper is to bridge the gap between the marketing and ESG literatures by conceptualizing "synthetic authenticity" and proposing a framework for managing risks in AI-driven ESG communication in the context of the EU AI Act and CSRD regulations. Drawing on recent research in AI marketing, ESG communication, brand authenticity, and AI governance, and integrating the European regulatory context, the paper develops a five-layer conceptual framework linking AI-enabled communication quality to stakeholder trust, corporate reputation, and competitive advantage. The framework incorporates governance moderators and non-linear feedback loops, arguing that trust built on synthetic authenticity is structurally fragile and that its detection triggers asymmetric reputational damage. The paper contributes original conceptualisation at the intersection of generative AI, sustainable branding, and stakeholder trust theory.

1. Introduction

Generative AI has moved rapidly from a peripheral marketing tool to a central mechanism of corporate communication, including ESG communication. Organisations now deploy large language models to draft sustainability reports, generate stakeholder-targeted ESG content, monitor reputational signals in real time, and personalise ESG messaging (Davenport et al., 2020; Dwivedi et al., 2023). The efficiency case is clear. What is less clear is whether this acceleration of ESG content production represents a genuine advance in corporate transparency or a new and more technically sophisticated form of reputational construction. This paper argues that existing research underestimates a specific risk that generative AI introduces into ESG communication. The literature on greenwashing documents intentional selective disclosure corporations choosing what to reveal and

what to conceal (Lyon & Montgomery, 2015; Seele & Gatti, 2017). Generative AI creates a distinct type of problem when ESG narratives are algorithmically optimised for credibility, emotional resonance, and stakeholder approval, potentially diverging from organisational reality not because of deliberate deception, but due to the inherent structural properties of the technology itself. This paper terms this phenomenon synthetic authenticity and positions it as the primary conceptual contribution.

This paper is motivated by three research gaps. First, the marketing and ESG communication literatures remain siloed. Frameworks that address AI in marketing do not engage with ESG accountability dynamics, and ESG communication research has not theorised the implications of generative AI at the level of mechanism. Second, synthetic authenticity as a category is not conceptualised in either literature it is conflated with greenwashing, obscuring its distinct causal logic. Third, the governance conditions that determine whether AI-assisted ESG communication builds or erodes stakeholder trust are absent from current models. The paper addresses these gaps through a critical conceptual synthesis and situates the analysis within the European regulatory context of the EU AI Act (European Parliament, 2024) and the Corporate Sustainability Reporting Directive – CSRD (European Commission, 2023a).

2. Methodology

The aim of this paper is to bridge the gap between the marketing and ESG literatures by conceptualizing "synthetic authenticity" and proposing a framework for managing risks in AI-driven ESG communication in the context of the EU AI Act and CSRD regulations. The methodology of the paper is based on an integrative literature review and critical conceptual synthesis. This approach is chosen to overcome the siloing of existing marketing and ESG literatures. The paper uses inductive reasoning to define the phenomenon of "synthetic authenticity" based on an analysis of the structural properties of generative AI. Subsequently, a five-layer framework for managing risks in AI-driven ESG communication is developed through conceptual modelling. The methodological approach also includes a contextual analysis of European regulation (EU AI Act and CSRD), which serves as a normative basis for defining the governance moderation mechanisms in the proposed model. The methodological process was further supported by Claude Sonnet 4.6 (Anthropic) and ChatGPT (OpenAI), which were used for idea generation, source discovery, technical refinement, and support in the conceptual modelling and visual representation of the proposed framework. The authors retained full responsibility for the final conceptual synthesis and content integrity.

3. ESG communication, AI, and the authenticity problem

3.1. ESG communication and stakeholder trust

Effective ESG communication does more than transmit information, it constructs and sustains the accountability relationship between an organisation and the stakeholder audiences on whose assessments its reputation depends (Du et al., 2010; Hawn & Ioannou, 2016). Stakeholder trust, in this context, is not a passive response to disclosure volume but an active evaluation of whether an organisation's stated commitments are credible, consistent, and grounded in observable conduct (Christensen et al., 2021; Velte, 2022). The literature consistently shows that perceived gaps between ESG claims and actual performance produce non-linear trust penalties, particularly when stakeholders feel they have been misled rather than merely disappointed (Lyon & Montgomery, 2015). What is less theorised is how generative AI changes the trust-formation dynamic. Prior research examined trust in ESG communication under human authorship, where the cognitive and resource limitations

of human communicators created natural constraints on ESG narrative production. Generative AI removes those constraints and in doing so, it changes the evidentiary basis on which trust is built. When AI systems can generate ESG narratives indistinguishable in style and apparent commitment from human-authored ones, the conventional trust signals, emotional authenticity, narrative coherence, communicative effort, become less reliable indicators of substantive performance.

3.2. Sustainable branding and the authenticity condition

Sustainable corporate branding connects ESG values with a company's brand identity in a way that stakeholders see as credible and trustworthy (Keller, 2020). However, the success of ESG branding depends heavily on perceived authenticity. Research shows that when stakeholders believe ESG communication is driven mainly by reputational goals rather than genuine commitment, trust decreases and scepticism increases (Morhart et al., 2015; Vredenburg et al., 2020). This makes authenticity a key limitation of AI-assisted ESG branding. Brand authenticity is shaped by credibility, integrity, consistency, and long-term commitment (Morhart et al., 2015). Importantly, authenticity is not something companies can simply claim; it depends on how stakeholders perceive organisational actions and values over time (Moulard et al., 2016). AI-generated ESG communication may appear authentic on the surface, even when it is not fully connected to the company's real practices.

This paper argues that it is important to distinguish between greenwashing and synthetic authenticity. Greenwashing involves intentionally misleading or exaggerating ESG performance. Synthetic authenticity is different. It occurs when AI-generated ESG communication sounds convincing and emotionally credible but is not genuinely supported by the organisation's actual ESG activities. This can happen even without deliberate deception because AI systems are often optimised for engagement and positive reactions rather than for accuracy or materiality. The distinction matters because greenwashing requires regulation and enforcement, while synthetic authenticity requires transparency and accountability in how AI systems are designed and used. Greenwashing can be seen as the result of a strategic decision (top-down), where the goal is to manipulate audience perceptions for reputational benefit. In synthetic authenticity, the culprit is an uncontrolled process, where a manager may use AI in good faith to streamline reporting, but the resulting product (narrative) will differ from reality precisely because of the AI's pursuit of a perfect and convincing tone.

3.3. Generative AI in ESG communication: Capabilities and governance gaps

Generative AI now encompasses automated CSRD and Global Reporting Initiative (GRI) aligned report drafting, AI-powered stakeholder sentiment analysis, dynamic ESG content personalisation, and real-time reputational monitoring (Davenport et al., 2020; Dwivedi et al., 2023; Floridi et al., 2020). In practice, organisations such as large financial institutions and FTSE 100 companies have begun integrating large language models into sustainability reporting workflows, reducing production timelines while maintaining the appearance of detailed disclosure. The efficiency gains are real, but the governance architecture required to ensure that AI-generated ESG content remains accurate, material, and substantively grounded in operational reality has not developed at a comparable pace.

The EU AI Act (European Parliament, 2024) introduces transparency and accountability requirements for AI systems used in contexts affecting public interests, requirements that plausibly extend to AI-assisted ESG communication given its materiality for investors, regulators, and civil society. The CSRD's double materiality standard (European Commission, 2023a; 2023b) requires organisations to disclose both financial and societal impact materiality, a standard that AI content systems must be explicitly configured to meet, not merely to approximate. The gap between regulatory expectations and current AI deployment practice in ESG communication is substantial and, this paper argues, underestimated in both the marketing and sustainability management literatures.

4. Benefits and risks of AI-Driven ESG communication

4.1. Strategic benefits

Three benefits of AI-assisted ESG communication are analytically significant not because they are novel, but because they relate to the trust-formation mechanisms theorised in the framework:

1. *Disclosure consistency*. Generative AI can enforce terminological and narrative coherence across multi-framework ESG reporting, CSRD, GRI, TCFD (Task Force on Climate-related Financial Disclosures), in ways that are difficult to achieve through human editorial processes at scale. While achieving such consistency is not a new goal and has traditionally been driven by manual review, generative AI acts as a "noise reduction tool" that systematically eliminates editorial inconsistencies and "noise" that typically arise in high-volume, human-driven reporting workflows. Inconsistent ESG disclosure is itself a trust-erosion signal, associated empirically with higher perceived greenwashing risk (Marquis et al., 2016). AI-assisted consistency thus has a direct, not merely an operational, relevance to stakeholder trust.
2. *Personalised stakeholder engagement*. AI-enabled segmentation allows the same substantive ESG data to be communicated in forms calibrated to investor, consumer, regulatory, and employee audiences. This is not a superficial customisation function: research in ESG communication suggests that comprehension and perceived relevance are key mediators of stakeholder trust formation (Puntoni et al., 2021; Christensen et al., 2021). The benefit of generative AI lies in the ability to automatically create a vast amount of unique communication outputs tailored to specific target groups while maintaining the integrity and relevance of the underlying ESG data. This real-time personalization capability allows organizations to have a more credible and effective dialogue with stakeholders, as it bridges the gap between the technical nature of ESG reporting and stakeholders' information needs.
3. *Internal governance support.* AI tools that cross-reference published ESG claims against internal operational data can serve a quasi-audit function, surfacing potential inconsistencies before disclosure. This internal verification mechanism, largely absent from the marketing and branding literature, positions AI as an ESG governance tool rather than a communication production system (Floridi et al., 2020). The benefit of this mechanism is protection against unintentional greenwashing, i.e. situations where differences between public claims and operational reality result from data complexity or reporting errors rather than intentional deception.

4.2. Risks: Synthetic authenticity, Greenwashing amplification, and reputational asymmetry

The main argument of this paper is that synthetic authenticity should be understood as a risk distinct from greenwashing. Treating these two concepts as equivalent creates confusion in both theory and practice. Greenwashing usually involves intentional misrepresentation. Organisations selectively present ESG information or exaggerate their sustainability performance to create a more positive image (Lyon & Montgomery, 2015; Seele & Gatti, 2017). Synthetic authenticity, however, does not always require deliberate deception. For example, a company may use a large language model trained on ESG reports from successful firms to generate its own sustainability communication. The resulting content can sound professional, convincing, and emotionally engaging while overstating the company's actual ESG ambitions or commitments. In this case, the AI is not intentionally "lying"; it is simply producing language that appears credible. The problem is that the communication may appear authentic without being genuinely grounded in organisational practices. The consequences for stakeholder trust can be even more serious than in traditional greenwashing cases. Research on AI-

mediated communication suggests that stakeholders expect higher levels of accountability when AI-generated content is involved (Hancock et al., 2020). If stakeholders realise that ESG communication was algorithmically designed to appear authentic rather than reflect genuine commitment, this can create a strong sense of betrayal. As a result, organisations may not only lose trust, but stakeholders may also begin questioning the credibility of previous ESG communication.

Another important risk is that AI can significantly amplify greenwashing. Traditional greenwashing is limited by human time and resources, whereas AI systems can generate and distribute large amounts of ESG content almost instantly and at very low cost. Without proper governance, this capacity for mass overstatement means that what stakeholders perceive as lying becomes a systemic risk, as misleading narratives are spread simultaneously across sustainability reports, investor presentations, social media, and marketing communication at near-zero marginal cost. This creates a level of reputational risk that is much larger than in conventional greenwashing and is still not fully addressed in existing risk management frameworks.

5. Framework for managing risks in AI-Driven ESG communication

The five-layer theoretical framework presented in Figure 1 explains how AI-supported ESG communication can contribute to competitive advantage when it is supported by strong governance and perceived as authentic by stakeholders. This framework also considers feedback effects and identifies synthetic authenticity at Layer 3 (L3) as the main risk area. The framework was derived through an integrative synthesis of AI governance standards (EU AI Act), strategic management theory - the Resource-Based View (RBV), and established ESG communication models to capture the full trajectory from technological input to market outcome.

Layer 1 (L1) includes the generative AI and Natural Language Processing (NLP) tools organisations use to create ESG communication. However, these tools cannot be viewed as neutral technologies. The EU AI Act's requirements for transparency and human oversight (European Parliament, 2024) directly affect how organisations can use AI without creating regulatory or reputational risks. Companies that focus only on efficiency and neglect governance increase the likelihood of synthetic authenticity emerging later in the process.

Layer 2 (L2) focuses on the quality of ESG communication, particularly whether AI-generated content meets standards of accuracy, completeness, and materiality. The CSRD's double materiality principle is especially important here (European Commission, 2023a). High AI capability alone does not guarantee high-quality ESG communication. Outcomes depend on data quality, external assurance, and whether AI systems prioritise material information or stakeholder approval. This tension is where synthetic authenticity begins to emerge.

Layer 3 (L3) is the key part of the framework because stakeholder trust depends not only on communication quality, but also on perceived authenticity. Stakeholders need to believe that ESG communication reflects real organisational practices rather than content designed to appear authentic. The paper argues that transparency about AI use is essential at this stage. Organisations that openly disclose AI involvement and demonstrate human oversight are more likely to maintain trust than those where AI use is revealed later. In some cases, transparency about AI-assisted communication may even strengthen stakeholder trust. The framework also includes two feedback loops. *A positive loop* exists where stronger stakeholder trust improves corporate reputation and ESG brand value and competitive advantage. At the same time, the framework highlights *a negative loop*: if stakeholders detect greenwashing or synthetic authenticity, reputational damage can quickly reduce competitive advantage and increase long-term scepticism. This suggests that the risks of synthetic authenticity failure may outweigh the efficiency benefits AI brings to ESG communication.

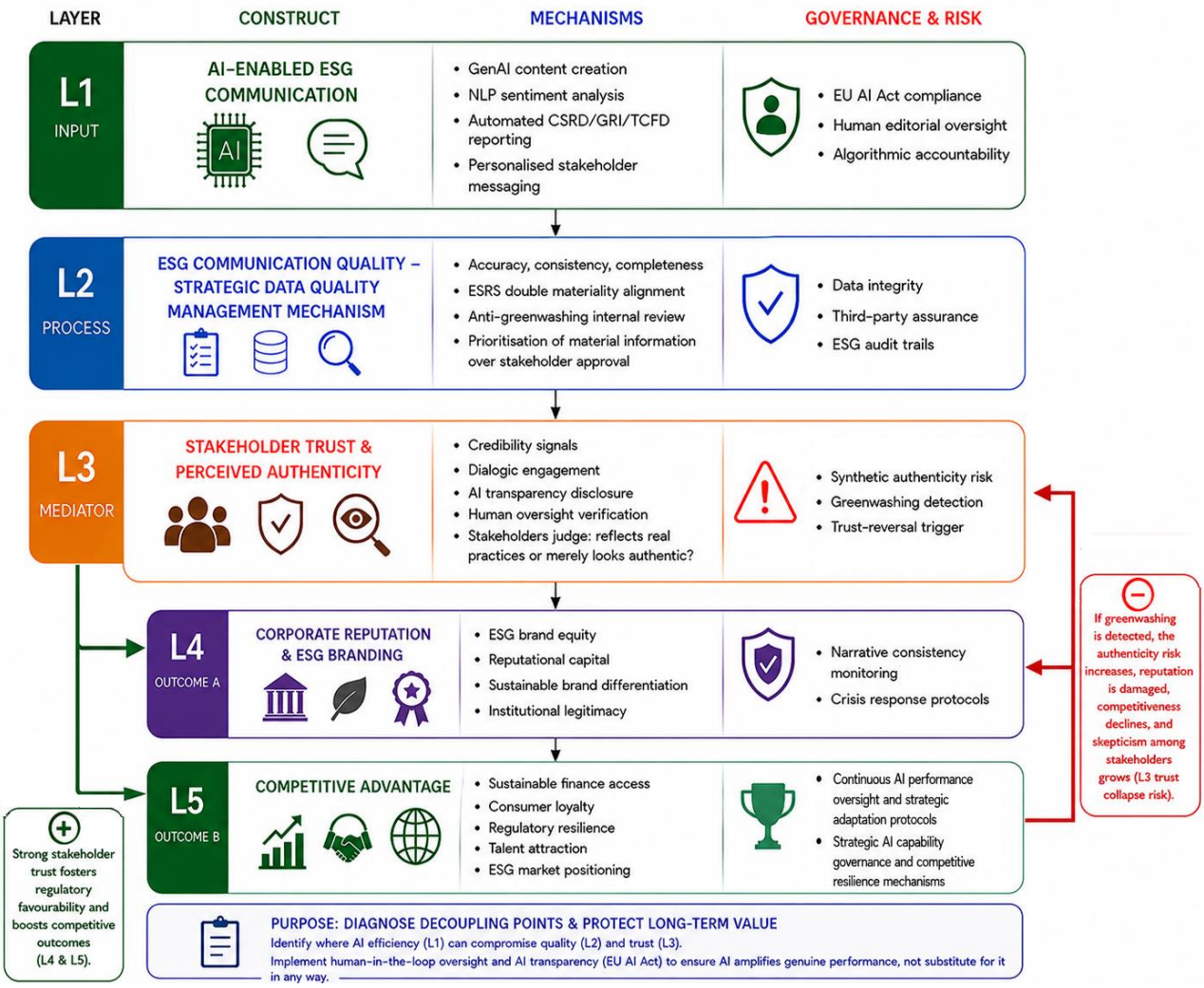


Figure 1. The Five-Layer Theoretical Framework for Managing Risks in AI-Driven ESG Communication

Source: authors' own conceptualisation

Layer 4 (L4) and Layer 5 (L5) represent the reputational and competitive outcomes of the framework's upstream dynamics, collectively constituting a joint value realization phase. Within the logic of the RBV theory, corporate reputation (L4) serves as a strategic intangible resource that transforms directly into measurable competitive advantages (L5) through sustained alignment between what the brand communicates and what the organisation delivers (Barney, 1991; Keller, 2013). Discussing these layers together underscores the theoretical necessity of their alignment: without reputational legitimacy (L4), it is impossible to secure a durable and sustainable market position (L5). The competitive advantages of strong ESG branding as sustainable finance access, consumer loyalty, regulatory resilience are real and well-documented (Eccles & Klimenko, 2019). However, the framework illustrates that these advantages accrue durably only when ESG communication is substantively grounded. Synthetic authenticity erodes this alignment, trading long-run competitive durability for short-run communicative efficiency, thus, AI-assisted communication must amplify genuine performance rather than substitute for it.

This model serves as a diagnostic tool for managers to identify "decoupling points" where AI efficiency (L1) may compromise substantive quality (L2) and stakeholder trust (L3). By surfacing the non-linear feedback loops between reputation (L4) and trust triggers, the framework provides a roadmap for implementing human-in-the-loop oversight and AI transparency disclosures required by

the EU AI Act. It helps organisations move beyond efficiency gains toward a governance-sensitive approach that protects long-term competitive value against the risks of synthetic authenticity.

6. Conclusion

This paper has advanced a conceptual argument with three interlocking components. First, it theorises synthetic authenticity as a risk category distinct from greenwashing, arising from the structural properties of generative AI rather than from deliberate strategic deception, and producing stronger trust-reversal dynamics because stakeholders perceive it as communicative manipulation rather than performance failure. Second, it argues that governance architecture including AI transparency, human editorial oversight, and compliance with the EU AI Act and CSRD is not an optional addition to AI-assisted ESG communication but the condition under which AI deployment produces durable reputational outcomes. Third, it proposes a non-linear framework that captures both the trust-building and trust-destroying dynamics of AI-enabled ESG communication, highlighting the asymmetry between efficiency gains and reputational risks. Taken together, these arguments suggest that organisations investing in generative AI for ESG communication without corresponding investment in governance infrastructure are not merely assuming compliance risk; they are creating the conditions for synthetic authenticity and the reputational reversal it can trigger. AI transparency, third-party ESG assurance, and human oversight of AI-generated sustainability content are therefore not costs to minimise but governance mechanisms that determine whether AI-assisted ESG branding creates or erodes competitive advantage.

The conceptual nature of this paper entails several limitations. Although the framework is theoretically grounded, its relationships require empirical validation across stakeholder groups, industries, and regulatory contexts. Future research should develop measures of synthetic authenticity perception, experimentally test the non-linear trust-reversal hypothesis, and examine how AI transparency disclosures vary across CSRD-regulated organisations. As generative AI becomes more deeply embedded in corporate communication, these questions will gain increasing practical relevance. While our framework is grounded by prior theory (e.g., Lyon & Montgomery, 2015; Morhart et al., 2015), future research is needed to determine whether stakeholders penalise "AI lies" more severely than intentional human deception, or vice versa.

Acknowledgement

This paper represents a partial output of the project VEGA No. 1/0733/25 "ESG reporting and examining the impact of achieving ESG criteria on the financial performance of companies in the Slovak Republic" and was supported by this project.

References

Barney, J. (1991). Firm resources and sustained competitive advantage. Journal of Management, 17(1), 99–120.

Christensen, L. T., Morsing, M., & Thyssen, O. (2021). Talk–action dynamics: Modalities of aspirational talk. Organization Studies, 42(3), 407–427. https://doi.org/10.1177/0170840619896380

Csikosova, A., Culkova, K., & Janoskova, M. (2016). Evaluation of quantitative indicators of marketing activities in the banking sector. Journal of Business Research, 69(11), 5028-5033. https://org.doi/10.1016/j.jbusres.2016.04.075

Davenport, T., Guha, A., Grewal, D., & Bressgott, T. (2020). How artificial intelligence will change the future of marketing. Journal of the Academy of Marketing Science, 48(1), 24–42. https://doi.org/10.1007/s11747-019-00696-0

Du, S., Bhattacharya, C. B., & Sen, S. (2010). Maximising business returns to corporate social responsibility (CSR): The role of CSR communication. International Journal of Management Reviews, 12(1), 8–19. https://doi.org/10.1111/j.1468-2370.2009.00276.x

Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., Baabdullah, A. M., Koohang, A., Raghavan, V., Ahuja, M., Albanna, H., Albashrawi, M. A., Al-Busaidi, A. S., Balakrishnan, J., Barlette, Y., Basu, S., Bose, I., Brooks, L., Buhalis, D., & Wright, R. (2023). 'So what if ChatGPT wrote it?' Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. International Journal of Information Management, 71, 102642. https://doi.org/10.1016/j.ijinfomgt.2023.102642

Eccles, R. G., & Klimenko, S. (2019). The investor revolution. Harvard Business Review, 97(3), 106–116.

European Commission. (2023a). Corporate Sustainability Reporting Directive (CSRD): Directive 2022/2464/EU. Official Journal of the European Union. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32022L2464

European Commission. (2023b). European Sustainability Reporting Standards (ESRS): Commission Delegated Regulation (EU) 2023/2772. Official Journal of the European Union. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ:L_202302772

European Parliament. (2024). Regulation (EU) 2024/1689 of the European Parliament and of the Council on artificial intelligence (AI Act). Official Journal of the European Union. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ:L_202401689

Floridi, L., Cowls, J., King, T. C., & Taddeo, M. (2020). How to design AI for social good: Seven essential factors. Science and Engineering Ethics, 26(3), 1771–1796. https://doi.org/10.1007/s11948-020-00213-5

Hancock, J. T., Naaman, M., & Levy, K. (2020). AI-mediated communication: Definition, research agenda, and ethical considerations. Journal of Computer-Mediated Communication, 25(1), 89–100. https://doi.org/10.1093/jcmc/zmz022

Hawn, O., & Ioannou, I. (2016). Mind the gap: The interplay between external and internal actions in the case of corporate social responsibility. Strategic Management Journal, 37(13), 2569–2588. https://doi.org/10.1002/smj.2464

Keller, K. L. (2013). Strategic brand management: Building, measuring, and managing brand equity (4th ed.). Pearson Education.

Keller, K. L. (2020). Consumer research insights on brands and branding: A JCR curation. Journal of Consumer Research, 46(5), 995–1001. https://doi.org/10.1093/jcr/ucz058

Lyon, T. P., & Montgomery, A. W. (2015). The means and end of greenwash. Organization & Environment, 28(2), 223–249. https://doi.org/10.1177/1086026615575332

Marquis, C., Toffel, M. W., & Zhou, Y. (2016). Scrutiny, norms, and selective disclosure: A global study of greenwashing. Organization Science, 27(2), 483–504. https://doi.org/10.1287/orsc.2015.1039

Morhart, F., Malär, L., Guèvremont, A., Girardin, F., & Grohmann, B. (2015). Brand authenticity: An integrative framework and measurement scale. Journal of Consumer Psychology, 25(2), 200–218. https://doi.org/10.1016/j.jcps.2014.11.006

Moulard, J. G., Raggio, R. D., & Folse, J. A. G. (2016). Brand authenticity: Testing the antecedents and outcomes of brand management's passion for its products. Psychology & Marketing, 33(6), 421–436. https://doi.org/10.1002/mar.20888

Puntoni, S., Reczek, R. W., Giesler, M., & Botti, S. (2021). Consumers and artificial intelligence: An experiential perspective. Journal of Marketing, 85(1), 131–151. https://doi.org/10.1177/0022242920953847

Seele, P., & Gatti, L. (2017). Greenwashing revisited: In search of a typology and accusation-based definition incorporating legitimacy strategies. Business Strategy and the Environment, 26(2), 239–252. https://doi.org/10.1002/bse.1912

Velte, P. (2022). Archival research on integrated reporting: A systematic review of main drivers and the impact of integrated reporting on firm value. Journal of Management and Governance, 26(3), 997–1031. https://doi.org/10.1007/s10997-021-09569-9

Vredenburg, J., Kapitan, S., Spry, A., & Kemper, J. A. (2020). Brands taking a stand: Authentic brand activism or woke washing? Journal of Public Policy & Marketing, 39(4), 444–460. https://doi.org/10.1177/0743915620947359

IDENTIFYING INNOVATION IN PUBLIC PROCUREMENT: A TEXT MINING ANALYSIS OF V4 COUNTRIES

Miroslava Barkóciová, Mariana Ivaničková

Bratislava University of Economics and Business, Slovakia
Faculty of Business Economics with seat in Kosice
miroslava.barkociova@euba.sk, mariana.ivanickova@euba.sk

DOI: 10.35011/IDIMT-2026-275

Keywords

CPV; public procurement; innovation; text mining; natural language processing; V4 countries.

Abstract

Public procurement represents an important policy instrument for supporting innovation, yet its measurement remains methodologically challenging due to the lack of standardized indicators. This study aims to identify innovation-related public procurement in the Visegrad (V4) countries and analyse its distribution across countries and sectors. The analysis is based on a large dataset of procurement notices obtained from the OpenTender platform. A rule-based text mining approach was applied to tender titles and descriptions using a multilingual keyword dictionary capturing technological, process and environmental innovation. Innovation was operationalised through both a broad and a strict definition to assess sensitivity to methodological choices. The results reveal substantial cross-country differences, with Poland exhibiting a significantly higher share of innovation-related procurement compared to Slovakia, the Czech Republic and Hungary. Sectoral analysis based on CPV categories shows that innovation-related procurement is concentrated in specific domains, particularly in construction, healthcare and information technology. Despite differences in overall levels, several CPV categories recur across all countries. The findings demonstrate the usefulness of text mining for analysing innovation in public procurement while highlighting the importance of definition choice and data characteristics in shaping results.

1. Theoretical background

Public procurement has increasingly been conceptualised not only as an administrative mechanism for acquiring goods, services and works, but also as a strategic policy instrument capable of shaping markets, supporting innovation and addressing broader socio-economic objectives. In this context, procurement data provide a valuable empirical basis for analysing how public demand is articulated across sectors and countries. The increasing digitisation of procurement procedures has resulted in large-scale datasets containing both structured variables, such as CPV codes, contract values and buyer characteristics, and unstructured textual information, including tender titles and descriptions.

Public procurement is globally recognized as the government activity most vulnerable to corruption (Abdou et al., 2022; Yanuarisa et al., 2025). The success of sustainable procurement depends heavily

on effective leadership, while inadequate management commitment serves as a primary barrier (Zaidi et al., 2019; Ershadi et al., 2021). However, leveraging these barriers by stakeholders can improve contract administration and reduce bottlenecks (Oyewobi & Jimoh, 2022). While advanced analytics and robotics modernize procurement and limit fraud, technology alone is not enough; success requires bridging finance, technology, and public administration to eliminate systemic flaws caused by funding gaps and insufficient training (Yanuarisa et al., 2025).

The innovative potential of procurement is amplified when combined with science policy instruments and supply-side or technology-push tools (Caravella & Crespi, 2020; Kundu et al., 2025). Despite this, a significant methodological gap remains regarding data collection and frameworks to evaluate the actual impact of these strategies (Kundu et al., 2025). Efforts toward technological sovereignty must be proactive, focusing on cultivating domestic capacities and global participation rather than mere defensiveness (Edler et al., 2023). Therefore, procurement discussions must increasingly consider supply chain resilience and alignment with broader industrial, environmental, and social goals (Edler et al., 2023; Kundu et al., 2025).

Textual elements in procurement notices contain analytically relevant signals; their linguistic and semantic characteristics can predict procurement outcomes (Acikalin et al., 2023), and text mining/NLP tools effectively extract information from unstructured datasets (Zhang et al., 2023). A related stream of research focuses on object classification using the EU's Common Procurement Vocabulary (CPV), demonstrating that tender descriptions can automatically generate CPV codes (Basile et al., 2023) or serve hierarchical machine learning problems (Moiraghi et al., 2024). These text-based methods are crucial in cross-country, multilingual settings, where deep learning supports document processing across European languages (Bednar et al., 2025). This is especially vital for the Visegrad countries, where notices span Slovak, Czech, Polish, Hungarian, and English, requiring an analytical approach sensitive to linguistic variation while maintaining cross-country comparability.

Within this framework, innovation in public procurement is identified through text mining of tender descriptions (e.g., digitalization, automation, sustainability). While CPV categories are not inherently innovative, certain sectors exhibit a higher frequency of innovation-related terms. This study combines text mining with CPV classification to identify innovation-oriented procurement notices in V4 countries and analyse their distribution. This approach reveals sectors where innovation repeatedly occurs, improving the understanding of innovation demand in procurement data.

2. Methodology

This study applies a quantitative empirical approach combined with text mining to analyse innovation patterns in V4 public procurement. The framework extracts textual information from procurement notices and links it to sectoral classifications using CPV categories.

2.1. Data

The analysis is based on publicly available procurement data obtained from the OpenTender platform, specifically from the dataset available at https://opentender.eu/all/download. This dataset integrates procurement data from multiple European sources, including national procurement systems and the European Tenders Electronic Daily (TED) database.

The study focuses on the V4 countries, namely Slovakia, Czech Republic, Poland and Hungary. The unit of analysis is an individual procurement notice (tender). The procurement data were downloaded from the OpenTender platform on 26. April 2026. The analysis covers procurement notices published

between 2009 and 2024. After data cleaning and filtering procedures, the final dataset contained 251,954 procurement notices from the four Visegrad countries.

From the original dataset, only a subset of variables relevant for the analysis was selected, including tender title (tender_title), tender description (tender_description), CPV classification (cpvcat or tender_mainCpv), year of publication (tender_year), and country of the contracting authority (buyer_country). The choice of the buyer country reflects the national context of public procurement policy, rather than the origin of suppliers.

It should be noted that OpenTender aggregates procurement data from multiple sources with potentially different levels of coverage across countries and years. Consequently, the dataset may not capture the full universe of procurement activities in each country. Differences in procurement volumes should therefore be interpreted with caution, as they may partly reflect variations in data availability and reporting practices.

2.2. Data preprocessing

Prior to analysis, the dataset was preprocessed for consistency. Irrelevant variables and columns with high missing values were removed, reducing the dataset to a minimal structure needed for text analysis and sectoral classification. Textual data were constructed by combining the tender title and description into a single lowercase field, cleaned of numbers, special symbols, and redundant whitespace. Additionally, CPV codes were standardized to their two-digit form, enabling meaningful higher-level sectoral comparison while avoiding category fragmentation.

2.3. Identification of innovation

Innovation in public procurement is not directly observable and therefore requires an indirect operationalisation. In this study, innovation is identified using a rule-based text mining approach based on a predefined multilingual keyword dictionary. The keyword dictionary was developed based on a review of the innovation procurement literature, policy documents related to innovation and digital transformation, and an iterative examination of procurement notice terminology within the analysed countries.

Table 1. Representative keyword categories

Category	**Examples of keywords**
Digitalisation	digitalisation, digital, e-government, online platform
Artificial intelligence and data analytics	artificial intelligence, machine learning, data mining, big data, predictive analytics
Automation and robotics	automation, robotic, robotisation, RPA
Smart technologies and IoT	smart, intelligent, IoT, Internet of Things, sensors
Software and information systems	software, application, information system, platform, cloud, cybersecurity
Modernisation*	innovation, modernisation, new solution, novel, pilot, experimental
Environmental innovation*	sustainability, ecological, green, renewable, low-emission, circular

Source: Authors' own elaboration based on the multilingual keyword dictionary developed for this study

*Included only in the broad definition of innovation. Equivalent terms in Slovak, Czech, Polish and Hungarian were also included to reflect the multilingual nature of the dataset.

Each procurement notice was classified using a binary variable (innovation_flag), where a value of 1 indicates the presence of at least one innovation-related keyword in the text, and 0 otherwise. This

approach allows for a scalable and transparent classification of large datasets without requiring manually labelled training data.

To enhance robustness, the analysis considers two alternative definitions of innovation. The first, broader definition includes general terms such as “modernisation” or “improvement”, capturing incremental innovation. The second, stricter definition focuses on advanced digital and technological concepts, including digitalisation, artificial intelligence, data analytics, automation, robotics, smart technologies and information systems.

To provide an initial assessment of classification validity, a random sample of 50 procurement notices identified as innovation-related was manually reviewed. Most sampled notices contained explicit references to innovation-related technologies, digitalisation, automation, sustainability, or related concepts. The results suggest that the rule-based approach provides a reasonable approximation of innovation-oriented procurement activities, although some degree of classification error cannot be excluded.

2.4. Analytical approach

The analytical part of the study is based on descriptive statistics and comparative analysis. First, the share of innovation-related procurement notices is calculated for each country, allowing for cross-country comparison within the V4 region. Second, innovation-related tenders are aggregated according to CPV categories to identify sectors with a higher concentration of innovation-related procurement.

Finally, a cross-country comparison identifies CPV categories that repeatedly exhibit innovation-related characteristics. Rather than assuming sectors are inherently innovative, this approach focuses on recurring patterns within specific categories. This combined method reveals sectoral and geographical trends, providing insights into how innovation-oriented demand is distributed across countries and sectors.

3. Results

The analysis identified substantial variation in the share of innovation-related procurement across the V4 countries, depending on the applied definition of innovation (see Table 1). Using the broader definition, which includes general innovation-related terms such as modernization and improvement, the highest share of innovation-related procurement was observed in Poland (78.5%). This was followed by Slovakia (25.9%) and Czech Republic (18.4%), while Hungary exhibited a significantly lower share (3.8%).

Table 2. Innovation share in public procurement by country (strict vs. broader definition)

Country	total_ tenders	Broader		Strict	
		innovative_tenders	innovation_share	innovative_tenders	innovation_share
PL	98 228	77 106	78.50%	76 310	77.69%
SK	32 972	8 541	25.90%	3 357	10.18%
CZ	65 471	12 020	18.36%	5 552	8.48%
HU	55 283	2 095	3.79%	1 929	3.49%

Source: based on OpenTender data

When applying the stricter definition of innovation, focusing on advanced technological and digital concepts, the overall levels of innovation-related procurement decreased across all countries.

However, the relative ordering remained consistent. Poland continued to show the highest share (77.7%), followed by Slovakia (10.2%), the Czech Republic (8.5%), and Hungary (3.5%).

Poland exhibits a notably higher share of innovation-related procurement notices than the other V4 countries under both definitions of innovation. Moreover, the difference between the broad and strict definitions is less than one percentage point, whereas substantially larger differences are observed in Slovakia and the Czech Republic. This finding may indicate a higher prevalence of explicit digital and technology-oriented terminology in Polish procurement documentation. However, differences in language usage, procurement reporting practices, and database coverage may also contribute to the observed pattern. Therefore, the Polish results should be interpreted with caution and warrant further investigation in future research.

Comparison between broad and strict definitions

Comparing the two approaches reveals substantial differences in measured innovation, particularly in Slovakia and the Czech Republic. In Slovakia, the share decreased from 25.9% under the broad definition to 10.2% under the strict definition; similarly, the Czech Republic declined from 18.4% to 8.5%. This suggests their procurement involves more general or incremental innovation rather than advanced technology. In contrast, Poland's share decreased only marginally from 78.5% to 77.7%, indicating a consistently high presence of innovation terminology. Hungary exhibited the lowest overall levels and smallest variation, with shares of 3.8% and 3.5%, respectively.

Cross-country differences

The results indicate pronounced cross-country differences in innovation-related procurement. Poland consistently demonstrates a substantially higher share than other V4 countries under both definitions. Slovakia and the Czech Republic occupy intermediate positions, showing high sensitivity to the applied definition, while Hungary shows consistently low levels regardless of the definition used. Overall, the findings suggest that measuring procurement innovation is highly dependent on the operational definition while revealing persistent differences across countries.

Distribution across CPV categories

The distribution of innovation-related procurement across CPV categories reveals distinct sectoral patterns within and across the V4 countries. Under the broader definition of innovation, several CPV categories consistently appear among the most frequent across all countries. Category 45 (construction work), 33 (medical equipment and pharmaceuticals), 38 (laboratory and precision equipment), 72 (IT services), and 48 (software packages) are prominently represented in multiple countries.

In the Czech Republic, the highest number of innovation-related tenders was observed in CPV category 45, followed by categories 33 and 38. A similar pattern can be observed in Slovakia, where category 45 dominates, followed by categories 42, 72, and 77. In Hungary, categories 33 and 45 are the most prominent, while in Poland, category 33 significantly dominates, followed by category 90 (environmental services) and category 77.

When applying the stricter definition of innovation, the overall distribution remains relatively stable, although some shifts in ranking can be observed. In the Czech Republic, category 45 remains dominant, while categories 38, 72, and 48 gain relative importance. In Slovakia, category 72 (IT services) becomes the most prominent, followed by category 45, indicating a stronger concentration of strictly defined innovation in digital sectors. Hungary shows a largely unchanged structure, with category 33 remaining dominant. In Poland, the distribution remains highly stable, with category 33 maintaining a substantially higher count compared to all other categories.

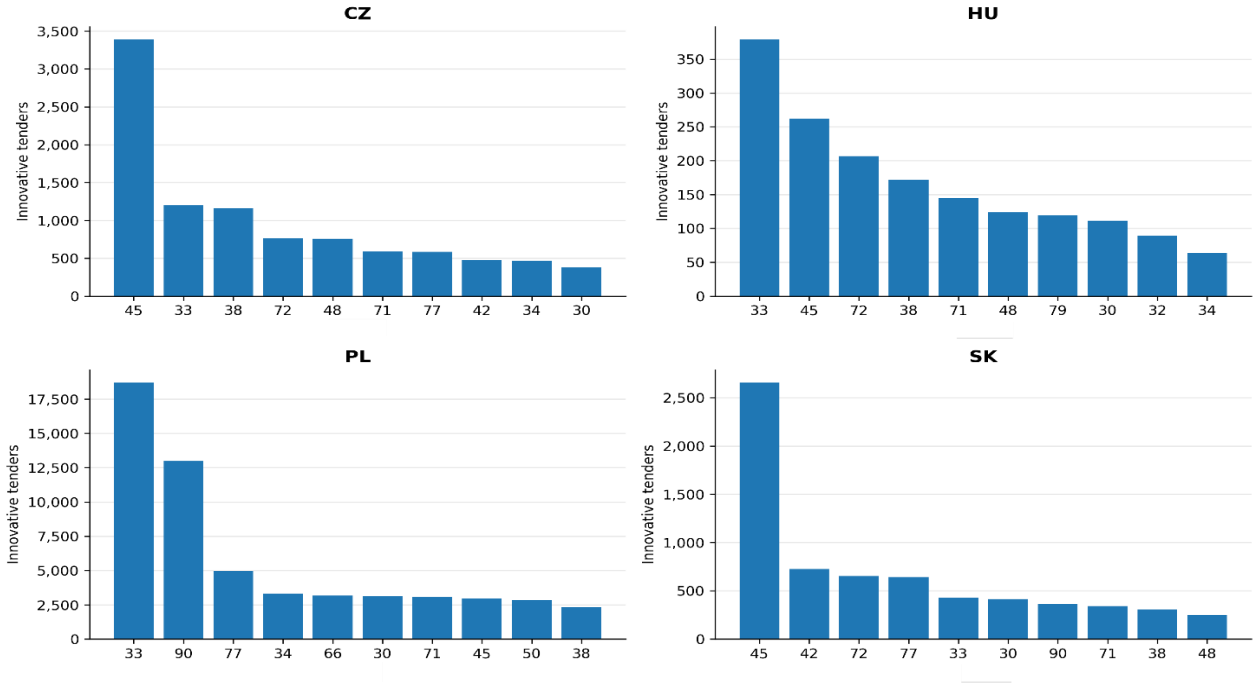


Figure 1. Top CPV categories of innovation-related procurement by country (broad definition)

Source: authors based on OpenTender data

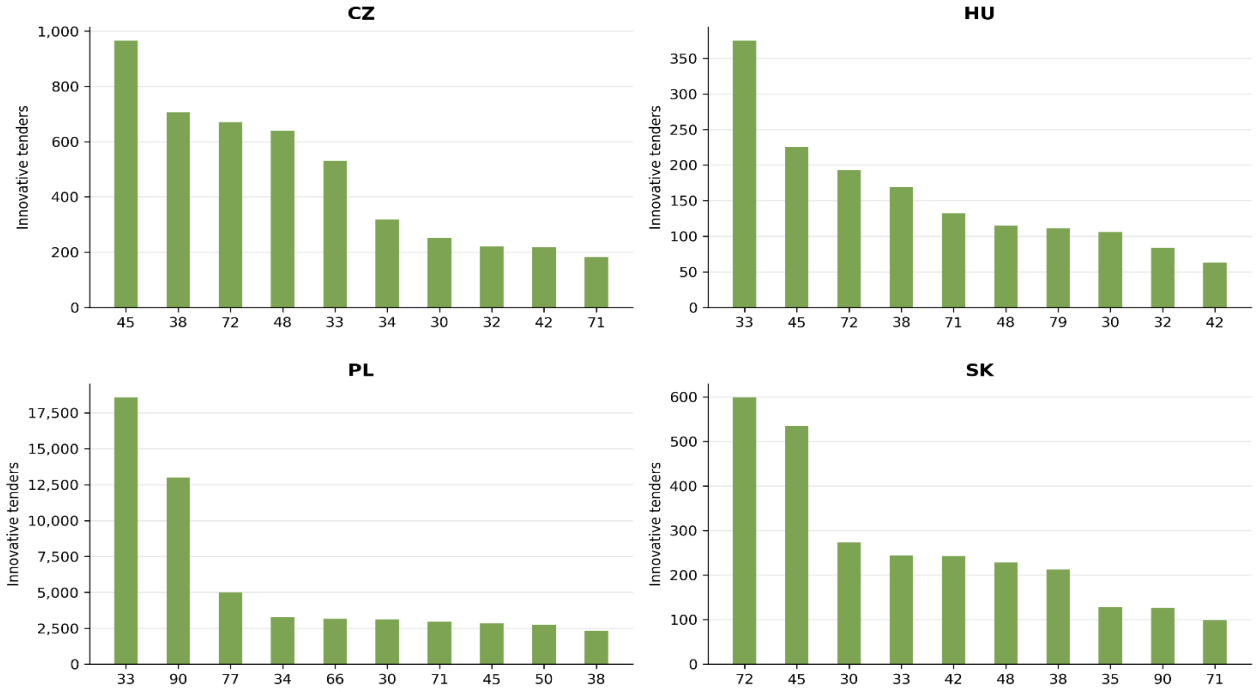


Figure 2. Top CPV categories of innovation-related procurement by country (strict definition)

Source: authors based on OpenTender data

Cross-country comparison of CPV categories

A cross-country comparison reveals that several CPV categories (33, 45, 72, 38, 48, and 30) are consistently present across all V4 countries under both definitions, suggesting that innovation-related

procurement clusters in specific sectors rather than being randomly distributed. Despite variations in the absolute number of tenders, the sectoral composition remains consistent. However, some country-specific patterns exist; for instance, categories 90 (environmental services) and 77 (agricultural services) are prominent in Poland and Slovakia but less pronounced in Hungary and the Czech Republic. Overall, the sectoral distribution exhibits both common and country-specific characteristics.

4. Discussion

The results of this study reveal substantial differences in the distribution of innovation-related public procurement across the V4 countries, as well as consistent sectoral patterns. These findings highlight both the importance of methodological choices in measuring innovation and the structural characteristics of public procurement systems in the analysed countries.

Poland shows a significantly higher share of innovation-related procurement than other V4 countries across all definitions. This may be due to Poland's larger market volume or differences in document formulation, where innovation terminology is more prevalent. As suggested by Acikalin et al. (2023), description wording influences analytical outcomes, meaning these differences may partly reflect linguistic practices rather than actual innovation.

The comparison between broad and strict definitions underscores how results depend on how innovation is operationalized. In Slovakia and the Czech Republic, a sharp decline under the stricter definition suggests procurement involves incremental improvements rather than advanced technological innovation, reflecting the known methodological challenge of measuring both forms. Conversely, Poland shows consistent innovation terminology even under strict criteria, while Hungary exhibits both low innovation levels and minimal variation between definitions, indicating limited innovation content or less explicit articulation in procurement texts.

The sectoral analysis based on CPV categories provides additional insights into the structure of innovation-related procurement. The repeated occurrence of categories such as 33 (medical equipment), 45 (construction), 72 (IT services), 38 (laboratory equipment), and 48 (software) across all countries suggests that innovation-related procurement tends to cluster in specific sectors. This finding is consistent with previous research indicating that certain domains, particularly information technology and healthcare, are more likely to involve innovation-driven procurement.

At the same time, results indicate CPV categories are not inherently innovative; instead, certain sectors simply exhibit a higher frequency of innovation-related characteristics. This distinction avoids overgeneralization and acknowledges the role of textual representation. The observed shift in Slovakia, where IT-related categories dominate under the stricter definition, illustrates this importance. While broader definitions capture a wider range of activities, stricter criteria highlight sectors driven by digital and technological transformation, directly impacting which sectors are identified as "innovation - intensive."

Despite the strengths of this approach, several limitations exist. Innovation identification relies on a rule-based keyword method sensitive to keyword selection, multilingual dataset variations, and terminology differences across languages. Furthermore, this method captures the presence of innovation-related terms rather than actual implementation or impact. As highlighted by Zhang et al. (2023) regarding NLP applications in procurement, extracting meaningful insights from unstructured data remains highly dependent on data quality and contextual interpretation.

Future research could address these limitations by incorporating machine learning approaches, such as supervised classification models or contextual embeddings, to improve the accuracy of innovation

detection. Additionally, combining textual analysis with structured data, such as contract values or supplier characteristics, could provide a more comprehensive understanding of innovation in public procurement. Overall, the findings contribute to the growing body of research on data-driven analysis of public procurement by demonstrating how text mining can be used to identify innovation-related patterns and compare them across countries and sectors. The results highlight both the potential and the limitations of text-based approaches in capturing complex phenomena such as innovation in public procurement.

Acknowledgement

This research was conducted within the framework of the project A-25-107/3020-19: Optimization of financial investments in innovation in industry using artificial intelligence. The results presented in this paper represent a partial output of the project.

References

Abdou, A., Basdevant, O., David-Barrett, E., & Fazekas, M. (2022). Assessing vulnerabilities to corruption in public procurement and their price impact. IMF Working Papers (094). https://doi.org/10.5089/9798400207884.001

Acikalin, U.U., Gordun, M.K., Kutlu, M., & Tas, B.K.O. (2023). How you describe procurement calls matters: Predicting outcome of public procurement using call descriptions. Natural Language Engineering, 30(6), 1255-1276. https://doi.org/10.1017/s135132492300030x

Basile, V., et al. (2023). Automatic generation of Common Procurement Vocabulary codes. Zenodo. https://doi.org/10.5281/zenodo.10007544

Bednar, P., et al. (2025). Deep learning methods for multilingual classification of business documents. IEEE International Symposium on Applied Machine Intelligence and Informatics. https://doi.org/10.1109/sami63904.2025.10883330

Caravella, S., & Crespi, F. (2021). The Role of Public Procurement as Innovation Lever: Evidence from Italian Manufacturing Firms. Economics of Innovation and New Technology, 30(7), 663-684. https://doi.org/10.1080/10438599.2020.1761591

Edler, J., Blind, K., Kroll, H., & Schubert, T. (2023). Technology sovereignty as an emerging frame for innovation policy. Defining rationales, ends and means. Research policy 52(6), 104765. https://doi.org/10.1016/j.respol.2023.104765

Ershadi, M., Jefferies, M., Davis, P., & Mojtahedi, M. (2021). Barriers to achieving sustainable construction project procurement in the private sector. Cleaner Engineering and Technology, 3, 100125. https://doi.org/10.1016/j.clet.2021.100125

Kundu, O., Uyarra, E., Ortega-Argiles, R., Tirado, M.M., Kitsos, T., & Yuan, P.Y. (2025). Impacts of policy-driven public procurement: a methodological review. Science and Public Policy, 52(1), 50–64. https://doi.org/10.1093/scipol/scae058

Moiraghi, F., Palmonari, M., Allavena, D., & Morando, F. (2024). Zero-shot Hierarchical classification on the Common Procurement Vocabulary taxonomy. In IEEE 48th Annual Computers, Software, and Applications Conference (pp. 273-278). https://doi.org/10.48550/arxiv.2405.09983

Oyewobi, L.O., & Jimoh R.A. (2022). Barriers to Adoption of Sustainable procurement in the Nigerian Public Construction Sector. Sustainability, 14(22), 14832. https://doi.org/10.3390/su142214832

Yanuarisa, Y., Irianto, G., Djamhuri, A., & Rusydi, M. K. (2025). Exploring the internal audit of public procurement governance: A systematic literature review. Cogent Business & Management, 12(1), 2485411. https://doi.org/10.1080/23311975.2025.2485411

Zaidi S. A. H. et al. (2019). Addressing the sustainable development through sustainable procurement: what factors resist the implementation of sustainable procurement in Pakistan? Socio-Economic Planning Sciences, 68, 100671. https://doi.org/10.1016/j.seps.2018.11.008

Zhang, X., et al. (2023). Mining healthcare procurement data using text mining and natural language processing: Reflection from an industrial project. arXiv preprint. https://doi.org/10.48550/arxiv.2301.03458

DETERMINANTS OF TAX RELIABILITY AMONG SLOVAK CONSTRUCTION SMES

Marian Brontvaj, Robert Verner

Bratislava University of Economics and Business, Slovakia
Faculty of Business Economics with seat in Kosice
marian.brontvaj@euba.sk, robert.verner@euba.sk

DOI: 10.35011/IDIMT-2026-283

Keywords

Tax reliability index; construction sector; small and medium-sized enterprises; ordered logit; Slovakia; financial performance.

Abstract

In 2022 the Slovak Financial Administration introduced a publicly disclosed Tax Reliability Index classifying every value-added-tax-registered taxpayer into one of four tiers. This paper investigates which firm-level characteristics explain tax reliability in a particularly heterogeneous and risk-prone segment of the economy, namely small and medium-sized construction completion and specialised construction enterprises. Using a hand-cleaned cross-sectional data set of 1,329 Slovak construction SMEs with annual revenues between EUR 1 million and EUR 15 million, drawn from the public business registry and the financial statements register, we estimate binary logit models, complemented by an ordered logit specification and an ordinary least squares benchmark on the three-tier ordinal recoding of tax reliability, on a battery of firm-level explanatory variables including profitability, size, age, ownership type, regional location and prior legal-risk events. We find that profitability and firm size are robust positive predictors of reliability, while the presence of historical payment orders and larger employee counts are associated with lower reliability in the bottom-tier margin. Marginal effects and odds-ratio transformations are reported to ease interpretation. The findings suggest that the new index captures genuine economic fundamentals and should be useful as a screening signal for business partners, lenders and public procurement authorities.

1. Introduction

Tax compliance in small and medium-sized enterprises (SMEs) is a persistent policy concern across Europe and particularly in the Central and Eastern European (CEE) region, where the value-added-tax (VAT) gap has historically exceeded the Union average. In Slovakia, the Financial Administration reported a VAT gap of around 14% as recently as 2021, with construction consistently identified as one of the industries most exposed to under-declaration, phoenix-firm behaviour and missing-trader fraud. Against this background, in January 2022 the Slovak Financial Administration launched a publicly disclosed Tax Reliability Index classifying every VAT-registered taxpayer into one of four tiers: highly reliable, reliable, less reliable, and unrated. The index combines information on the

timeliness and completeness of tax filings, the regularity of payments, and the outcomes of past tax audits, and is intended both as an ex-post supervisory tool and as an ex-ante market signal.

The introduction of a disclosed compliance score offers an unusually clean setting in which to test which firm-level fundamentals translate into observable tax behaviour. The construction sector is a particularly informative laboratory because it combines low barriers to entry, high labour-intensity, thin margins, and well-documented exposure to informal work and chain-type VAT fraud. Within the sector, completion and specialised construction works are of particular interest: dominated by micro and small enterprises, they operate at the downstream end of the value chain and are the most common providers of services to private households, where invoicing discipline is weakest.

This paper contributes in three ways. First, we focus on a single, well-defined sub-population of 1,329 Slovak construction SMEs with annual revenue between EUR 1 million and EUR 15 million, controlling for unobserved sectoral heterogeneity while isolating the genuine SME segment. Second, we use the four-tier Tax Reliability Index both as a binary and as an ordinal outcome, distinguishing general reliability from top-tier reliability. Third, we combine accounting fundamentals with legal-risk flags (payment orders, bankruptcies, executions) and governance variables (ownership type, age, regional location), giving a fuller view of the drivers of compliance than the accounting-only models usual in prior work.

The contribution is also relevant for supply-chain risk management and procurement analytics. Construction projects are delivered through long subcontracting chains in which a general contractor relies on dozens of specialised SME suppliers, and the tax reliability of each link is a material risk factor: VAT non-compliance by a subcontractor exposes downstream firms to joint-and-several liability for unpaid VAT under Slovak law, and tax-driven insolvency of a key subcontractor can disrupt project delivery. A model that explains the disclosed Tax Reliability Index from observable firm characteristics is therefore directly usable in supplier risk-scoring systems, vendor pre-screening pipelines, and procurement decision-support tools that rank bidders by ex-ante default and compliance risk.

2. Literature review

The firm-level determinants of tax compliance have been studied from several angles. Allingham and Sandmo (1972) provide the classic economics-of-crime framework, in which firms weigh the gains from under-reporting against the probability and severity of detection, later extended with behavioural, institutional and reputational dimensions (Alm, 2019; Kirchler, 2007; Hashimzade et al., 2013). Randomised audit experiments such as Kleven et al. (2011) show that third-party-reported income leaves little scope for evasion, whereas self-reported items are systematically under-declared, and Pomeranz (2015) demonstrates that the self-enforcing structure of VAT generates strong deterrence even without audits. A recurring finding is that profitability and firm size are positively associated with compliance, because larger and more profitable firms face higher reputational and audit costs, while liquidity-constrained firms are more likely to defer tax payments (Nurtegin, 2008; Joulfaian, 2009; Devereux et al., 2014).

In the CEE context, Koumpias and Martinez-Vazquez (2019) document that transparency initiatives such as publicly disclosed taxpayer registers materially improve reported compliance, consistent with the field-experimental evidence of Fellner et al. (2013) on Austrian TV-licence fees. Joulfaian (2009) and Putnins and Sauka (2015) provide complementary evidence from transition economies: non-compliance concentrates in smaller and less profitable firms and is driven as much by the perceived quality of tax administration as by the tax burden. Recent work by Belas et al. (2020) and Dvorsky et

al. (2021) on Slovak and Czech SMEs highlights the role of financial performance and external financing in the risk profile of regional SMEs.

Studies that directly exploit the 2022 Slovak Tax Reliability Index in econometric work are still scarce, as it has been publicly disclosed only since the 2022 reform. Kmetova et al. (2023) provide an early descriptive analysis of the index, while Mihokova et al. (2018) use factor analysis to categorise Slovak corporate taxpayers and find that financial-performance variables carry most of the discriminative power. The closest accounting-based reference points are studies of Slovak construction firms: Horvathova and Mokrisova (2018) examine bankruptcy prediction and argue that ratios of profitability, liquidity and indebtedness are the most discriminative predictors, a variable selection that provides a natural benchmark here.

3. Hypotheses

The Slovak Tax Reliability Index applies to VAT-registered legal persons, each classified on the basis of an internal scoring rule into one of four tiers: highly reliable, reliable, less reliable, and unrated. Highly reliable taxpayers enjoy procedural benefits such as shorter audit windows and preferential VAT-refund timing, while less reliable taxpayers face more frequent audits and longer refund processing. The unrated tier covers taxpayers for whom insufficient rating-relevant information is available, typically very young or dormant firms.

Building on the literature and the specifics of Slovak construction SMEs, we formulate four hypotheses. First, because more profitable firms enjoy higher operating cash flows, they are better positioned to settle tax liabilities on time and fund a compliance function; we expect profitability to be a positive predictor of reliability. Second, larger firms, by revenue and by employment, face higher reputational and procedural costs of non-compliance, are more visible to auditors, and maintain more formalised accounting; we expect firm size to be positively associated with the highest tier. Third, firms that have accumulated payment orders, executions or a bankruptcy history reveal weak payment discipline likely to spill over into the tax domain; we expect such legal-risk flags to reduce reliability. Finally, ownership type and regional location should play a secondary but significant role, with foreign-owned firms and firms in more developed regions more likely to reach the higher tiers.

4. Data and methodology

The primary data source is a commercial business-information database consolidating information from the Slovak Commercial Registry. The initial extract covers 1,329 Slovak enterprises active in two four-digit SK-NACE construction classes with reported annual revenues between EUR 1 million and EUR 15 million. This band isolates the genuine SME segment: micro firms below EUR 1 million are excluded for unreliable, discontinuous reporting, while firms above EUR 15 million belong to the large-enterprise category with different governance. The reference date of the extract is 19 April 2026. The sample is dominated by the 43.39 completion-and-finishing class (865 firms, 65.1%), with the remaining 464 firms (34.9%) in the 43.99 specialised-activities class. Geographically, Bratislava (28.4%), Zilina (19.0%) and Nitra (12.3%) contribute the largest shares.

The dependent variable is the Tax Reliability Index, used in three operationalisations. The first is a binary logit distinguishing the bottom margin (reliable or highly reliable versus less-reliable firms); the second a binary logit distinguishing the top margin (highly reliable versus reliable, restricted to the rated-and-compliant subsample). Grouping the two upper tiers in the first specification is deliberate: for supply-chain and procurement decisions the relevant distinction is between compliant

and non-compliant counterparties, while the difference between reliable and highly reliable is largely procedural. The third operationalisation recodes the index as an ordinal variable taking the values 1 (less reliable), 2 (reliable) and 3 (highly reliable), estimated both by an ordered logit and by ordinary least squares with heteroskedasticity-robust standard errors. The unrated tier is excluded from all regressions because the underlying classification rule is not directly comparable to the reliability tiers. The main specification is a binary logit model of the form:

$$Pr(Reliable_i=1) = \Lambda(\alpha + \beta' X_i + \gamma' R_i + \delta' G_i)$$

where $\Lambda(\cdot)$ is the logistic CDF, X_i the vector of accounting fundamentals, R_i the legal-risk flags, and G_i the governance and location variables. Explanatory variables fall into three blocks. Accounting fundamentals include the natural logarithm of revenue, the profit margin (winsorised at the 1st and 99th percentiles), the midpoint employee count from the published size bucket, and firm age in years. Legal-risk flags are binary indicators for at least one payment order, for any recorded bankruptcy or restructuring event, and for an executed loss in the latest reported financial year. Governance and location variables include foreign ownership, a NACE dummy for the completion class, and regional dummies. Standard errors are heteroskedasticity-robust. For ease of interpretation, we transform the logit coefficients into odds ratios and average marginal effects, reported for the headline regressors in the text.

5. Results

Table 1 reports descriptive statistics by tax reliability tier. The pattern is economically intuitive: highly reliable firms have an average profit margin of 7.5%, compared with 1.2% for reliable, 1.3% for less reliable, and −1.0% for unrated firms. The loss-making share rises monotonically from 3.3% in the top tier to 27.8% in the unrated tier. Highly reliable firms are also the largest by employee count within the rated-and-compliant group, but the less-reliable tier is paradoxically bigger on average (16.4 employees), driven by a few larger firms with thin margins.

Table 1. Descriptive statistics by tax reliability tier

Variable	Highly reliable	Reliable	Less reliable	Unrated
Observations	301	581	224	223
Mean revenue (EUR thousand)	3,033	2,242	2,577	2,455
Mean profit margin (%)	7.5	1.2	1.3	−1.0
Share of loss-makers (%)	3.3	14.1	17.4	27.8
Mean employees (midpoint)	13.8	8.6	16.4	9.0
Mean age (years)	15.1	13.1	13.3	15.3
Share with payment order (%)	9.3	8.4	13.4	18.4
Share foreign-owned (%)	10.3	7.6	32.6	37.2

Source: authors (Slovak Registry of Financial Statements and the Financial Administration disclosure of the Tax Reliability Index)

An unexpected pattern is that foreign ownership is over-represented in the less reliable and unrated tiers rather than in the top tier, consistent with foreign-owned firms often being dormant subsidiaries of international groups whose main operations are booked elsewhere. Table 2 reports the two logit specifications and the OLS benchmark on the 1 to 3 ordinal recoding. All specifications include regional fixed effects.

Table 2. Determinants of tax reliability. Logit and OLS estimates

Variable	(1) Logit Rel. vs Less	(2) Logit Highly vs Rel.	(3) OLS Ordinal 1–3
Log revenue	−0.166 (0.199)	+0.705*** (0.168)	+0.084** (0.040)
Profit margin (%, winsorised)	+0.039** (0.019)	+0.315*** (0.044)	+0.043*** (0.003)
Employees	−0.013*** (0.005)	+0.029*** (0.008)	+0.000 (0.002)
Age (years)	+0.031** (0.015)	+0.005 (0.013)	+0.006** (0.003)
Payment order (0/1)	−0.602* (0.331)	+0.002 (0.299)	−0.078 (0.072)
Loss in latest year (0/1)	+0.506 (0.462)	+0.616 (0.442)	+0.189*** (0.064)
Foreign-owned (0/1)	+0.157 (0.396)	+0.477 (0.323)	+0.116* (0.068)
Completion NACE (0/1)	−0.240 (0.233)	−0.168 (0.185)	−0.045 (0.039)
Region FE	Yes	Yes	Yes
Observations	946	843	946
McFadden / adj. R^2	0.041	0.290	0.164

Source: authors

Three findings stand out. First, profitability is the most consistent predictor of tax reliability. A one percentage-point increase in the profit margin raises the odds of being reliable by about 4% in column (1) and about 37% in column (2). Second, size effects are non-monotonic: employee count has a negative coefficient in column (1) but a strongly positive one in column (2). The labour-intensity channel must be distinguished from the revenue-based size channel. Once a SME crosses from less reliable to reliable, its employee count becomes a positive predictor of climbing to the top tier, so Hypothesis 2 is supported only in the high-tier margin. Third, the payment-order flag is significantly negative in column (1), reducing the odds of reliability by about 45% and the predicted probability of being reliable rather than less reliable by about 6 percentage points. This supports Hypothesis 3. The bankruptcy and loss flags are not individually significant. Hypothesis 4 receives only partial support: foreign ownership is positively signed throughout but significant only in the OLS benchmark.

Results are qualitatively unchanged when the sample is restricted to firms with 2024 or 2025 financials (593 and 410 observations), when standard errors are clustered regionally, when an ordered logit or probit is estimated on the three-tier recoding, and when the ratio of payment orders to revenue replaces the binary flag. The profitability and size coefficients are stable, and the non-monotonic employee effect survives all robustness tests.

6. Discussion

The evidence supports the view that the new index captures genuine economic fundamentals rather than only procedural filing behaviour. Profitability and size are standard proxies for the capacity to finance timely tax payments and staff a compliance function, and both are strongly positively associated with the classification. The payment-order flag is also telling: firms accumulating civil-law evidence of weak payment discipline towards private counterparties also tend to accumulate administrative evidence of the same towards the Financial Administration, a cross-validation that echoes Schneider and Enste (2013) on the contagion of informal practices across contractual relationships.

The non-monotonicity of the employee-count coefficient qualifies the view that larger firms are unambiguously more compliant. Among rated construction SMEs, several relatively large firms with thin margins appear in the less-reliable tier, consistent with the sector having pockets of semi-informal labour in which larger payrolls coexist with weaker tax-payment discipline. This echoes the size paradox of Klucnikov et al. (2022) and Williams and Horodnic (2017) on CEE labour markets, where envelope-wage practices concentrate in medium-sized labour-intensive firms whose margins do not comfortably accommodate full social-security contributions. Disentangling the payroll and profitability channels is a natural direction for further research, drawing on Putnins and Sauka (2015).

A further dimension is the interpretation of the unrated tier, which a naive view would treat as missing observations. The descriptive table shows, however, that unrated firms have the highest share of loss-makers (27.8%), of foreign ownership (37.2%) and of payment orders (18.4%). The tier is nonetheless heterogeneous, mixing very young firms, dormant foreign-owned subsidiaries and firms in reorganisation alongside firms with weak fundamentals, so the statistics describe the average rather than any individual firm. On average it contains a non-trivial share of weak-fundamental firms whose tax filings are dormant or disputed, in line with Torgler and Schneider (2009), who argue that residual classification tiers act as de facto warning labels. External users should therefore treat the tier as a flag for further due diligence rather than a direct compliance verdict.

Three caveats are in order. First, the index is updated infrequently and our outcome variable is therefore measured at a single point in time; a panel version would require re-scraping the Financial Administration website at several dates. Second, the employee count is reported in buckets and the midpoint imputation introduces classical measurement error that attenuates the estimated coefficient (Hansen et al., 2020). Third, the sample restricts attention to two four-digit NACE classes and to a narrow revenue band, and external validity to other construction activities remains to be established.

7. Conclusion

Using a cross-sectional sample of 1,329 Slovak construction SMEs with annual revenue between EUR 1 million and EUR 15 million (958 in the active regression sample), this paper investigates the firm-level determinants of the Slovak Tax Reliability Index introduced in 2022. Profitability and firm size are the most robust predictors of reliability, and civil-law payment orders significantly reduce the odds of being classified as reliable, while foreign ownership and regional location play secondary roles. The index therefore appears to summarise meaningful economic information and could be usefully incorporated, alongside standard financial ratios, in the credit-scoring models used by banks and in public procurement vendor selection. The relevance also extends to supply-chain risk management, where general contractors face joint-and-several VAT liability for the unpaid taxes of their subcontractors.

The most original finding is the non-monotonic role of employment: simply being a larger employer does not push a Slovak construction SME into the top tier, as it is the combination of size with profitability that matters. This warns against using employee count as a standalone compliance proxy. Future work should extend the analysis to the full construction sector, exploit repeated disclosures of the index to construct a panel, and test for the feedback effect of tier upgrades on firms subsequent revenue growth.

References

Allingham, M. G., & Sandmo, A. (1972). Income tax evasion: A theoretical analysis. Journal of Public Economics, 1(3–4), 323–338.

Alm, J. (2019). What motivates tax compliance? Journal of Economic Surveys, 33(2), 353–388.

Belas, J., Smrcka, L., Gavurova, B., & Dvorsky, J. (2020). The impact of social and economic factors in the credit risk management of SMEs. Technological and Economic Development of Economy, 24(3), 1215–1230.

Devereux, M. P., Liu, L., & Loretz, S. (2014). The elasticity of corporate taxable income: New evidence from UK tax records. American Economic Journal: Economic Policy, 6(2), 19–53.

Dvorsky, J., Petrakova, Z., Cera, G., & Folvarcna, A. (2021). Important factors for the entrepreneurship in central Europe. Innovative Marketing, 17(1), 61–73.

Fellner, G., Sausgruber, R., & Traxler, C. (2013). Testing enforcement strategies in the field: Threat, moral appeal and social information. Journal of the European Economic Association, 11(3), 634–660.

Hansen, B., Miller, S. V., & Weber, C. (2020). Federalism, partial prohibition, and cross-border sales: Evidence from recreational marijuana. Journal of Public Economics, 187, 104159.

Hashimzade, N., Myles, G. D., & Tran-Nam, B. (2013). Applications of behavioural economics to tax evasion. Journal of Economic Surveys, 27(5), 941–977.

Horvathova, J., & Mokrisova, M. (2018). Risk of bankruptcy, its determinants and models. Risks, 6(4), 117.

Joulfaian, D. (2009). Bribes and business tax evasion. European Journal of Comparative Economics, 6(2), 227–244.

Kirchler, E. (2007). The economic psychology of tax behaviour. Cambridge: Cambridge University Press.

Kleven, H. J., Knudsen, M. B., Kreiner, C. T., Pedersen, S., & Saez, E. (2011). Unwilling or unable to cheat? Evidence from a tax audit experiment in Denmark. Econometrica, 79(3), 651–692.

Klucnikov, A., Civelek, M., Krajcik, V., & Ondrejmiskova, I. (2022). Innovative regional development of the structurally disadvantaged industrial region by the means of the local currency. Acta Montanistica Slovaca, 27(1), 176–189.

Kmetova, O., Frenakova, M., & Krsekova, M. (2023). Tax reliability index in Slovakia: A descriptive analysis. Ekonomika a manažment, 20(2), 15–29.

Koumpias, A. M., & Martinez-Vazquez, J. (2019). The impact of media campaigns on tax filing: Quasi-experimental evidence from Pakistan. Journal of Asian Economics, 63, 33–43.

Mihokova, L., Andrejovska, A., & Martinkova, S. (2018). Categorization of corporate taxpayers in Slovakia: A factor analysis approach. Journal of Applied Economic Sciences, 13(7), 1896–1910.

Nurtegin, K. D. (2008). Determinants of business tax compliance. B.E. Journal of Economic Analysis & Policy, 8(1), 18.

Pomeranz, D. (2015). No taxation without information: Deterrence and self-enforcement in the value added tax. American Economic Review, 105(8), 2539–2569.

Putnins, T. J., & Sauka, A. (2015). Measuring the shadow economy using company managers. Journal of Comparative Economics, 43(2), 471–490.

Schneider, F., & Enste, D. H. (2013). The shadow economy: An international survey (2nd ed.). Cambridge: Cambridge University Press.

Torgler, B., & Schneider, F. (2009). The impact of tax morale and institutional quality on the shadow economy. Journal of Economic Psychology, 30(2), 228–245.

Williams, C. C., & Horodnic, I. A. (2017). Evaluating the illegal employer practice of under-reporting employees salaries. British Journal of Industrial Relations, 55(1), 83–111.

THE TRUST PHENOMENON IN AI-DRIVEN MARKETING: ESG AS A MITIGATING TOOL

Mária Dolná, Magdaléna Freňáková, Jozefína Hvastová

Bratislava University of Economics and Business, Slovakia
Faculty of Business Economics with seat in Kosice
maria.dolna@euba.sk, magdalena.frenakova@euba.sk, jozefina.hvastova@euba.sk

DOI: 10.35011/IDIMT-2026-291

Keywords

AI-driven marketing; consumer trust; ESG; transparency; explainability; algorithmic accountability.

Abstract

As artificial intelligence (AI) increasingly governs consumer interactions in digital marketing, a critical trust deficit has emerged between consumers and the opaque algorithmic systems that drive hyper-personalization. This paper investigates the multidimensional nature of this trust gap, asking how organizations can restore and sustain consumer trust in AI-driven marketing. Using a Systematic Literature Review following PRISMA 2020 guidelines, this paper synthesises research published between 2022 and 2026 from Scopus, Web of Science, ScienceDirect, and related databases. The review identifies and ranks ten trust-related phenomena, including the Personalization-Privacy Paradox, Black Box Problem, Algorithmic Bias, AI Washing, and AI Booing, through a phenomena matrix organised by prevalence and strategic significance. The findings suggest that Environmental, Social, and Governance (ESG) criteria provide a robust mitigating framework for reducing algorithmic opacity and rebuilding relational trust. The paper concludes that sustainable AI marketing requires organizations to treat trust as a strategic resource and empower consumers as stakeholders rather than passive targets of algorithmic surveillance.

1. Introduction

AI integration has enabled unprecedented hyper-personalization, but its rapid expansion has outpaced ethical governance, creating a structural trust deficit. This erosion of trust, driven by inaccessible decision logic, represents a strategic liability for organizations in regulated markets. This paper utilizes a systematic literature review to categorize these phenomena and proposes an ESG-based framework to restore algorithmic accountability. The primary objective of this paper is to investigate how organizations can enhance trust and transparency in AI-driven marketing strategies to improve customer engagement and loyalty. Specifically, the paper aims to *analyse the foundations of trust, identify and rank trust-related phenomena, formulate strategic responses to mitigate these phenomena,* and *incorporate ESG as a unique mitigating tool*.

AI integration has transformed marketing from broadcasting to hyper-personalized engagement (Chingono et al., 2026; Das et al., 2026). As AI governs interactions, an *algorithmic accountability gap* has emerged, eroding consumer trust (Borimnejad & Borimnejad, 2026; Narayan, 2025; Hari & Bibyana, 2025; Mollay et al., 2026). Trust depends on the perceived competence, transparency,

reliability, and ethical standing of automated systems rather than just brand reputation (Hari & Bibyana, 2025; Mary et al., 2026). This trust is relational, emerging from alignment with human expectations (Bach et al., 2024; Solovjov, 2025). Notably, 72% of consumers are more likely to trust brands that disclose the use of AI-generated materials (Morgan Stanley, 2024). A significant barrier to trust is the *Black Box Nature* of AI algorithms, which breeds skepticism regarding manipulation and surveillance (Mary et al., 2026; Chingono et al., 2026; Ozturkcan & Bozdağ, 2025; Waqas & Qadri, 2025). Scholars emphasize *Explainable AI (XAI)* to make algorithmic logic intelligible and foster accountability (Mary et al., 2026; Kumar & Dayal, 2025; Schnake et al., 2025; Hari & Bibyana, 2025; Chaudhary, 2024). This addresses the *Personalization-Privacy Paradox*, where consumers desire tailored experiences but reject invasive data collection (Adobe, 2023; Kumar & Dayal, 2025; Huda et al., 2024). Researchers advocate for *Personalization with Permission* through explicit consent and user-friendly privacy controls (Narayan, 2025). These trust dynamics are analysed through *Privacy Calculus Theory*, where consumers weigh personalization benefits against risks of privacy invasion and data misuse (Waqas & Qadri, 2025; Narayan, 2025; Chingono et al., 2026). Emerging threats such as *AI Washing*, exaggerating capabilities, fuel *AI Booing*, or public backlash (Ozturkcan & Bozdağ, 2025). Anthropomorphizing AI can enhance trust and emotional attachment, but it risks triggering the *Uncanny Valley* if transparency regarding its non-human nature is lacking (Wang et al., 2025; Mollay et al., 2026; Charles et al., 2025). Furthermore, a critical research gap exists in formulating a *holistic theory of algorithmic trust* (Chingono et al., 2026; Das et al., 2026), as current literature remains fragmented and Western-centric, overlooking enthusiastic AI adoption in Eastern markets (Das et al., 2026; Hari & Bibyana, 2025; Borimnejad & Borimnejad, 2026). Consequently, longitudinal and cross-cultural studies are essential to understand how trust dynamics evolve in global regulatory environments (Ozturkcan & Bozdağ, 2025; Solovjov, 2025).

2. Methodology

This paper adopts a Systematic Literature Review design following PRISMA 2020 guidelines to identify and synthesise research on AI-driven personalization and consumer trust. The search was conducted across Scopus, Web of Science, ScienceDirect, IEEE Xplore, EBSCOhost, and Google Scholar using a Boolean strategy: (“Artificial Intelligence” OR “AI”) AND (“Marketing”) AND (“Trust” OR “Transparency” OR “Ethics”). Publications were restricted to the 2022–2026 timeframe to focus on the generative AI wave, with selection involving title, abstract, and full-text assessments. Data were synthesised using a phenomena matrix, ranking prevalence based on citation frequency and recurring density across the literature. NotebookLM and Grammarly assisted in organization and language refinement, with all suggestions critically reviewed by the authors.

3. Results

AI integration enables hyper-personalized engagement at scale, but the shift toward autonomous systems has created a critical trust bottleneck. To address the fragmented academic landscape, this paper utilizes a phenomena matrix for synthesis.

3.1. Phenomena affecting consumer trust in AI-driven marketing

Based on the reviewed literature, Table 1 summarizes the main trust-related phenomena associated with AI-driven marketing and ranks them according to their prevalence in prior research.

Table 1. Matrix of trust-related phenomena in AI-driven marketing

Phenomenon	Authors	Description
1. Personalization, Privacy Paradox	(Adobe, 2023; Cloarec et al., 2024; Akter et al., 2022; Hussain, 2025; Chingono et al., 2026; Gupta et al., 2025; Chen, 2023)	The fundamental tension where consumers desire the benefits of hyper-tailored experiences but fear the losses of privacy, surveillance, and invasive data collection.
2. Black Box Problem, Algorithmic Opacity	(Mostafa, 2023; Marey et al., 2024; Bhatnagar, 2024; Chaudhary, 2024)	The phenomenon where algorithms process data in ways that are impenetrable even to experts, making the reasoning behind predictions or actions obscure to the user.
3. Algorithmic Bias, Discrimination	(Akter et al., 2022; Kumar & Dayal, 2025; Rini et al., 2024; Kordzadeh & Ghasemeghaei, 2022)	The systemic reproduction or amplification of existing social biases (racial, etc.) in marketing outcomes due to biased historical training data or flawed model architectures.
4. Uncanny Valley Effect	(Charles et al., 2025; Wang et al., 2025; Greilich, 2025)	The discomfort triggered when anthropomorphic (human-like) AI entities become too real but remain subtly artificial, causing perceptions of eeriness.
5. AI Booing	(Ozturkcan & Bozdağ, 2025; Chintalapati & Pandey, 2022)	The public backlash triggered by perceived AI failures, ethical lapses, bias, or surveillance concerns that lead to a so-called Cycle of AI Mistrust.
6. Filter Bubbles, Ideological Isolation	(Areeb et al., 2023; Rini et al., 2024; Querci et al., 2022)	An information environment where algorithms prioritize engagement-driven content that reinforces existing beliefs, thereby restricting exposure to diverse viewpoints.
7. AI Washing	(Ozturkcan & Bozdağ, 2025; Elhajjar & Itani, 2025)	The deliberate or negligent exaggeration of a system's AI capabilities, presenting rule-based systems as autonomous, to signal innovation without substantive backing.
8. Algorithmic Aversion	(Kumar & Dayal, 2025)	A psychological resistance where consumers distrust AI more after an error than they would a human, often leading to total disengagement from automated tools.
9. Ersatzing AI	(Madhavaram & Appan, 2024)	Artificial but inferior intelligence that makes marketing entities less capable due to inherent biases, errors, or significant gaps in knowledge.
10. Ethics Washing	(Munn, 2023; Ozturkcan & Bozdağ, 2025)	A practice where firms use superficial ethical principles or symbolic guidelines as a facade to mask structural concerns or avoid meaningful accountability.

Source: authors

3.2. Mitigating trust-related phenomena in AI-driven marketing

Mitigating the identified trust-related phenomena requires coordinated efforts across system design, institutional governance, and human-centred communication. Table 2 links each phenomenon identified in the literature with strategic responses aimed at strengthening consumer trust while ensuring that AI supports rather than replaces human judgment.

Table 2. Strategic responses to trust-related phenomena in AI-driven marketing

Phenomenon and authors	Strategic responses
Personalization-Privacy Paradox (Narayan, 2025; Chingono et al., 2026; Ozturkcan & Bozdağ, 2025; Hussain, 2025)	Adopt Permission-Based Frameworks: Organizations should move away from covert data exploitation and institutionalize personalization where consumers must actively opt-in.
Personalization-Privacy Paradox (Gartner, 2023; Chingono et al., 2026; Apple, 2023; Mollay et al., 2026)	Implement Privacy-by-Design: Trust is built when privacy is embedded into the initial design of marketing systems rather than added as a post-hoc compliance measure.
Personalization-Privacy Paradox (McKinsey, 2022; Alhitmi et al., 2024)	Utilize Privacy-Enhancing Technologies (PETs): Firms should invest in AI for data anonymization and federated learning to balance relevance with security.
Personalization-Privacy Paradox (Crowley & Standish, 2025; Kumar & Dayal, 2025)	Dynamic Privacy Dashboards: Providing users with interactive, real-time dashboards allows them to track gathered data, adjust settings, and revoke consent.
Black Box Problem/Algorithmic Opacity (Mostafa, 2023; Marey et al., 2024; Bhatnagar, 2024; Mary et al., 2026)	Implement Explainable AI (XAI): Using techniques like SHAP and LIME allows marketers to provide human-readable rationales for automated decisions.
Black Box Problem/Algorithmic Opacity (Jonker, 2024; Liao & Vaughan, 2024; Solovjov, 2025)	Adopt Human-Centred Transparency: Transparency should be a communication strategy using simple logic messages (e.g., "Because you watched...") rather than a technical code dump.
Black Box Problem/Algorithmic Opacity (Solovjov, 2025)	Establish Algorithmic Traceability: Use frameworks like audit trails to document a model's data lineage, version history, and bias test results.
Algorithmic Bias and Discrimination (Khamayseh et al., 2026; Solovjov, 2025; Hart, 2024)	Conduct Monthly Bias Audits: Employ ethics committees to run systematic scans simulating different demographic inputs to detect disparities in targeting.
Algorithmic Bias and Discrimination (Akter et al., 2022; Alshehadeh et al., 2026; Ozturkcan & Bozdağ, 2025)	Diverse Dataset Curation: Training data must be rigorously checked for diversity and inclusivity across racial, gender, and socioeconomic lines.
Algorithmic Bias and Discrimination (Khamayseh et al., 2026; Alshehadeh et al., 2026; Podevin et al., 2026)	Implement Human-in-the-Loop (HITL) Oversight: Combine AI efficiency with mandatory human confirmation for high-stakes decisions to prevent biased patterns.
AI Washing and AI Booing (Ozturkcan & Bozdağ, 2025)	Empirical Validation of Claims: Brands must ground AI claims in practical, validated applications with substantiated evidence rather than vague hype.
AI Washing and AI Booing (Ismatullaev & Kim, 2024; Ozturkcan & Bozdağ, 2025)	Calibrate Trust and Manage Expectations: Communicate both the capabilities and specific limitations of AI systems to avoid cycles of inflated expectations and backlash.
AI Washing and AI Booing (United Nations, 2025; Ozturkcan & Bozdağ, 2025)	Stakeholder Collaboration: Engage in multi-stakeholder dialogues to establish standardized industry definitions for legitimate AI innovation.
Filter Bubbles/Ideological Isolation (Areeb et al., 2023; Rini et al., 2024)	Content Diversity: Adjust algorithmic objectives to prioritize novelty and the inclusion of diverse viewpoints alongside relevance to historical behaviour.
Filter Bubbles/Ideological Isolation (Querci et al., 2022; Kumar & Dayal, 2025)	User Preference Customization: Platforms should allow users to manually escape feedback loops by offering easy-to-use settings to reset or modify preferences.

Phenomenon and authors	Strategic responses
Algorithmic Aversion (Kumar & Dayal, 2025)	Adopt Hybrid Intelligence Models: Use AI to augment rather than substitute human judgment to preserve the human touch and reduce consumer resistance.
Algorithmic Aversion (Hart, 2024; Solovjov, 2025)	Honest Error Disclosure: Admitting to algorithmic bugs and detailing steps taken to fix them is more effective at restoring trust than ignoring failures.
Ersatzing AI and Ethics Washing (Chingono et al., 2026; Madhavaram & Appan, 2024)	Institutionalize Multidisciplinary Governance: Legal, technical, and ethical professionals should collaborate throughout the AI lifecycle to embed principles into system architecture.
Ersatzing AI and Ethics Washing (Khamayseh et al., 2026; Solovjov, 2025)	Mandatory Staff Training: Ethical AI literacy training ensures marketing teams understand the broad social implications of the technology they deploy.

Source: authors

3.3. ESG as a mitigating tool for trust-related phenomena in AI-driven marketing

ESG criteria are increasingly recognized as an important framework for building stakeholder trust and enhancing the transparency of corporate activities. By integrating ESG principles into business processes and decision-making, organizations can increase the credibility of their activities, promote transparency, and create long-term value. As shown in Table 3, the environmental, social, and governance pillars provide several mechanisms for mitigating trust-related phenomena associated with AI-driven marketing practices.

Table 3. ESG-based mitigation mechanisms

Pillar	Mitigation mechanisms
Environmental (E) (Ozturkcan & Bozdağ, 2025)	Addresses the phenomenon of AI Washing by requiring technological claims to be anchored in measurable and verifiable environmental data. The concept of evidence-based sustainability requires companies to disclose the real-world impacts of operating AI systems, such as the carbon footprint and CO_2 emissions associated with training and operating large language models.
Social (S) (Podevin et al., 2026; Khamayseh et al., 2026)	This pillar mitigates Algorithmic Bias and the loss of human connection to prevent consumer resistance. By institutionalizing social accountability and responsible innovation, companies ensure algorithms don't reinforce biases. Maintaining a human touch through Green HRM (Human Resources Management) and hybrid intelligence is vital for social sustainability. Using AI to augment rather than replace human judgment preserves the empathy and creativity required for ethical decisions.
Governance (G) (Qasaimeh et al., 2026; Narayan, 2025)	Governance addresses the Black Box Problem through sustainable audits that increase transparency and traceability. By integrating ESG factors, companies help stakeholders understand AI decision-making. Furthermore, ethical governance resolves the Personalization-Privacy Paradox by treating trust as a strategic resource. Shifting from data exploitation to trust-based exchanges empowers consumers as stakeholders rather than surveillance targets.

Source: authors

4. Conclusion

This paper investigates how organizations can restore consumer trust in AI-driven marketing by identifying trust-related phenomena and proposing ESG as a mitigating tool. A systematic review identified ten critical phenomena mapped against organizational responses. The Personalization-

Privacy Paradox and the Black Box Problem emerged as primary challenges, requiring permission-based personalization, Explainable AI, and algorithmic traceability. The ESG framework addresses these by grounding claims in sustainability, promoting bias auditing, and institutionalizing transparency through governance mechanisms. Moving from algorithmic opacity to relational trust requires embedding ethical governance into system architecture rather than treating it as post-hoc compliance. This approach transforms consumers into empowered stakeholders, fostering durable brand loyalty. Future research should prioritize longitudinal and cross-cultural studies examining trust dynamics and ESG effectiveness in diverse regulatory contexts, particularly non-Western markets.

Acknowledgement

This contribution is part of the project A-25-107/3020-12 “Challenges in using artificial intelligence in economic and financial education: A multi-level educational framework“, and project VEGA No. 1/0733/25 “ESG reporting and examining the impact of achieving ESG criteria on the financial performance of companies in the Slovak Republic”.

References

Adobe. (2023). State of personalization in marketing. https://www.adobe.com

Akter, S., Dwivedi, Y. K., Sajib, S., Biswas, K., Bandara, R. J., & Michael, K. (2022). Algorithmic bias in machine learning-based marketing models. Journal of Business Research, 144, 201–216. https://doi.org/10.1016/j.jbusres.2022.01.083

Alhitmi, H. K., Mardiah, A., Al-Sulaiti, K. I., & Abbas, J. (2024). Data security and privacy concerns of AI-driven marketing in the context of economics and business field: An exploration into possible solutions. Cogent Business & Management, 11(1), 1–9. https://doi.org/10.1080/23311975.2024.2393743

Alshehadeh, A. R., El-Qirem, I. A., & Elrefae, G. A. G. A. (Eds.). (2026). Artificial intelligence in business: Applications and ethical considerations in the digital age. Springer Nature. https://link.springer.com/book/10.1007/978-3-031-96622-4

Apple. (2023). App tracking transparency: Enhancing user control. https://www.apple.com

Areeb, Q. M., Nadeem, M., Sohail, S. S., Imam, R., Doctor, F., Himeur, Y., Hussain, A., & Amira, A. (2023). Filter bubbles in recommender systems: Fact or fallacy—A systematic review. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 13(6), e1512. https://doi.org/10.1002/widm.1512

Bach, T. A., Khan, A., Hallock, H., Beltrão, G., & Sousa, S. (2024). A systematic literature review of user trust in AI-enabled systems: An HCI perspective. International Journal of Human–Computer Interaction, 40(5), 1251–1266. https://doi.org/10.1080/10447318.2022.2138826

Bhatnagar, N. (2024). Decoding closed box models with LIME (Local Interpretable Model-Agnostic Explanations). Medium.

Borimnejad, H., & Borimnejad, V. (2026). Emerging trends, challenges and research opportunities in artificial intelligence applications in marketing. Discover Artificial Intelligence, 6(9), 1–15. https://doi.org/10.1007/s44163-025-00705-y

Charles, V., Rana, N. P., & Jain, V. (2025). Marketing with anthropomorphised AI: Insights from consumers. Journal of Consumer Marketing, 42(4), 410-415. https://doi.org/10.1108/JCM-03-2025-7704

Chaudhary, G. (2024). Unveiling the black box: Bringing algorithmic transparency to AI. Masaryk University Journal of Law and Technology, 18(1), 93-122. https://doi.org/10.5817/MUJLT2024-1-4

Chen, J. (2023). AI personalization and consumer trust: Navigating the paradox of personalization and privacy in digital marketing. Journal of Behavioral and Experimental Economics, 107, 102108. https://doi.org/10.1016/j.chb.2024.108549

Chintalapati, S., & Pandey, S. K. (2022). Artificial intelligence in marketing: A systematic literature review. International Journal of Market Research, 64(1), 38-68. https://doi.org/10.1177/14707853211018428

Chingono, A. T. A., Chakweza, C. T., Kanyongo, R. S., & Tlou, R. (2026). AI-driven personalization in digital marketing: Balancing innovation and consumer privacy. International Journal of Computer Applications, 187(75), 68–85. https://doi.org/10.5120/ijca2026926297

Cloarec, J., Meyer-Waarden, L., & Munzel, A. (2024). Transformative privacy calculus: Conceptualizing the personalization-privacy paradox on social media. Journal of the Academy of Marketing Science, 52(3), 678–699. https://doi.org/10.1002/mar.21998

Crowley, J., & Standish, J. (2025). The empowered consumer. Accenture.

Das, A., Mishra, S., & Mayaluri, Z. L. (2026). Business intelligence, AI-driven personalization, and ethical customer segmentation in e-commerce: A systematic review of consumer trust and digital marketing. SN Business & Economics, 6(84), 1–29. https://doi.org/10.1007/s43546-026-01079-8

Elhajjar, S., & Itani, O. S. (2025). AI washing: A conceptual exploration. AMS Review. 15. https://doi.org/10.1007/s13162-025-00323-y

Gartner. (2023). AI and GDPR compliance predictions. https://www.gartner.com

Greilich, A., Bremser, K. & Wüst, K. (2025). Consumer Response to Anthropomorphism of Text-Based AI Chatbots: A Systematic Literature Review and Future Research Directions. International Journal of Consumer Studies 49(5), e70108. https://doi.org/10.1111/ijcs.70108.

Gupta, S., Pansari, A., & Kumar, V. (2025). All roads lead to Rome: The impact of personalization strategies on consumer engagement and trust. Journal of the Academy of Marketing Science, 53(2), 347–368. https://doi.org/10.1007/s11747-025-01063-w

Hari, K., & Bibyana, J. (2025). A cross-cultural framework for algorithmic trust: How data transparency mechanisms, 6(1), 727–743.

Hart, R. (2024). Clearview AI—controversial facial recognition firm—fined $33 million for 'illegal database'. Forbes. https://www.forbes.com/sites/roberthart/2024/09/03/clearview-ai-controversial-facial

Huda, M., Awaludin, A., & Siregar, H. K. (2024). Legal challenges in regulating artificial intelligence: A comparative study of privacy and data protection laws. International Journal of Social and Human, 1(2), 116–125. https://doi.org/https://dx.doi.org/10.59613/g8dc9v94

Hussain, S. R. (2025). The personalization paradox: When AI stops helping and starts overwhelming. SSRN.

Ismatullaev, U. V. U., & Kim, S. H. (2024). Review of the factors affecting acceptance of AI-infused systems. Human Factors, 66(1), 126–144. https://doi.org/https://doi.org/10.1177/00187208211064707

Jonker, A. (2024). What is AI transparency? IBM. https://www.ibm.com/think/topics/ai-transparency

Khamayseh, A. I. A., Yusoff, W. F. B. W., & Khudari, M. (2026). Immersive realities in e-marketing: Building consumer trust through ethical AR/VR experiences. In A. R. Alshehadeh, I. A. El-Qirem, & G. A. G. A. Elrefae (Eds.), Artificial Intelligence in Business: Applications and Ethical Considerations in the Digital Age (pp. 3–11). Springer Nature. https://doi.org/10.1007/978-3-031-96631-6_1

Kordzadeh, N., & Ghasemghaei, M. (2022). Algorithmic bias: Review, synthesis, and future research directions. European Journal of Information Systems, 31(3), 388–409. https://doi.org/10.1080/0960085X.2021.1927212

Kumar, A., & Dayal, M. (2025). AI-powered marketing: A content analysis of bias, transparency, and consumer trust. International Journal of Contemporary Research in Multidisciplinary, 4(3), 146–151.

Liao, Q. V., & Vaughan, J. W. (2024). AI transparency in the age of LLMs: A human-centered research roadmap. Harvard Data Science Review (Special Issue 5). https://doi.org/https://doi.org/10.1162/99608f92.8036d03b

Madhavaram, S., & Appan, R. (2024). AI for marketing: Enabler? Engager? Ersatz? Journal of the Academy of Marketing Science, 52(5), 1344–1351. https://doi.org/10.1007/s13162-024-00293-7

Marey, A., Arjmand, P., Alerab, A. D. S., Eslami, M. J., Saad, A. M., Sanchez, N., & Umair, M. (2024). Explainability, transparency and black box challenges of AI in radiology: Impact on patient care in cardiovascular radiology. Egyptian Journal of Radiology and Nuclear Medicine, 55(1), 183. https://doi.org/10.1186/s43055-024-01356-2

Mary, R., Esther, P., Bens, W., Ikhmat, A., & Lanre, A. (2026). Explainable AI in personalized marketing: Implications for trust and transparency in the tech industry. Procedia Computer Science.

McKinsey & Company. (2022). The state of AI in 2022—and a half decade in review. McKinsey Global Institute. https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai-in-2022-and-a-half-decade-in-review

Mollay, M., Sharma, D., Anawade, P., & Parlikar, C. (2026). The Impact of Privacy Regulations on Digital Marketing Practices: A Descriptive Study. In Senjyu, T., So-In, C., Joshi, A. (eds.) Smart Trends in Computing and Communications. SmartCom 2025. Lecture Notes in Networks and Systems, 1464, 81–90. Springer, Singapore. https://doi.org/10.1007/978-981-96-7520-3_8

Morgan Stanley. (2024). Generative AI: Labor & economy. Morgan Stanley Research.

Mostafa, S. (2023). Cracking the code: The black box problem of AI. ScaDS.AI. https://scads.ai/blog/2023/cracking-the-code-the-black-box-problem-of-ai/

Munn, L. (2023). The uselessness of AI ethics. AI and Society, 38(4), 1433–1443. https://doi.org/10.1007/s43681-022-00209-w

Narayan, R. (2025). Personalization with permission: Redefining consumer trust in AI-driven marketing models. International Journal of Economic Social Science and Management Law, 6(3), 14–25. https://ijeml.com/journal/index.php/ijeml/article/view/9

Ozturkcan, S., & Bozdağ, A. A. (2025). Responsible AI in marketing: AI booing and AI washing cycle of AI mistrust. International Journal of Market Research, 67(6), 696–722. https://doi.org/10.1177/14707853251379285

Podevin, D., Paulus, M., Wilhelm, A., & Hellbrück, N. (2026). Responsible Marketing in the Age of AI: Ethical Reflection for Companies Between Innovation and Responsibility. In Emrouznejad, A., Hammoudi, S., Wijnhoven, F. (eds) Smart Business Technologies. ICSBT 2025. Communications in Computer and Information Science, 2666, 216–228. Springer, Cham. https://doi.org/10.1007/978-3-032-08614-3_13

Rini, A., Wandrial, S., Lutfi, L., Jaya, I., & Satrionugroho, B. (2024). Data-Driven Marketing: Harnessing Artificial Intelligence to Personalize Customer Experience and Enhance Engagement. Join: Journal of Social Science. 1. 282-295. https://doi.org/10.59613/akx6j040

Qasaimeh, G., Al Khattab, A. N., Alrabba, H., Al Hassan, H. I., & Al-Shayea, Q. (2026). The impact of big data analytics on sustainable auditing in Jordanian commercial banks. In A. R. Alshehadeh, I. A. El-Qirem, & G. A. G. A. Elrefae (Eds.), Artificial Intelligence in Business: Applications and Ethical Considerations in the Digital Age (pp. 117–128). Springer Nature. https://doi.org/10.1007/978-3-031-96631-6_12

Querci, I., Barbarossa, C., Romani, S., & Ricotta, F. (2022). Explaining how algorithms work reduces consumers' concerns regarding the collection of personal data and promotes AI technology adoption. Psychology & Marketing, 39(10), 1874–1891. https://doi.org/10.1002/mar.21705

Schnake, T., Jafari, F. R., Lederer, J., Xiong, P., Nakajima, S., Gugler, S., Kloft, M., & Müller, K. R. (2025). Towards symbolic XAI – explanation through human understandable logical relationships between features. Information Fusion, 115, 102923. https://doi.org/10.1016/j.inffus.2024.102923

Solovjov, E. (2025). The impact of AI-generated advertisement on consumer trust, loyalty and emotional connection to a brand: Examining the moderating role of transparency, brand familiarity and the mediating role of perceived authenticity [Master's thesis, Universidade Católica Portuguesa]. Veritati Repository. https://repositorio.ucp.pt/bitstreams/75d95a8d-ae86-4974-8ad9-d15329fa8aed/download

United Nations. (2025). UN AI actions. AI for Good. https://aiforgood.itu.int/about-us/un-ai-actions/

Wang, Y., Sauka, K., & Situmeang, F. B. I. (2025). Anthropomorphism and transparency interplay on consumer behaviour in generative AI-driven marketing communication. Journal of Consumer Marketing, 42(4), 512–536. https://doi.org/10.1108/JCM-04-2024-6806

Waqas, M., & Qadri, U. A. (2025). When consumer autonomy is compromised: Adopting a dual-type view to understand AI-driven marketing and its impact on young consumers' ethical behavior. Journal of Business Ethics. https://doi.org/10.1007/s10551-025-06155-x

HETEROGENEITY OF REGISTERED SOCIAL ENTERPRISES IN PUBLIC PROCUREMENT: A DESCRIPTIVE ANALYSIS ON SLOVAK DATA

Abraham Gocze, Gabriel Demeter

Technical University of Kosice, Slovakia
Faculty of Economics
abraham.gocze@tuke.sk, gabriel.demeter@tuke.sk

DOI: 10.35011/IDIMT-2026-299

Keywords

CPV categories; social enterprises; public procurement; social procurement; procurement heterogeneity; competition intensity; savings; bidder behaviour; descriptive data analysis.

Abstract

This article examines heterogeneity in the behaviour and procurement outcomes of Registered Social Enterprises (RSEs) in Slovak public procurement using data from the Electronic Contracting System (EKS). Building on earlier descriptive work that documented low overall participation and strong concentration of RSE activity, the present study shifts attention from incidence to heterogeneity. It compares contracts with RSE participation and contracts without RSE participation and then explores how procurement outcomes vary across procurement categories. The analysis focuses on standard observable variables, especially savings, number of bidders, estimated contract value, and category structure based on CPV codes. The results suggest that RSE-related procurement should not be treated as a uniform segment. The observed patterns point to meaningful variation across procurement categories and firms, including differences in competition intensity, savings, portfolio breadth, and concentration of successful outcomes. The article remains descriptive and uses these findings to motivate later research using deeper analytical methods.

1. Introduction

Public procurement is increasingly understood not only as an administrative purchasing function, but also as a policy instrument through which public buyers may pursue wider economic, social, and environmental goals. In this broader setting, social enterprises are relevant because they combine commercial activity with a declared social mission and are often associated with employment creation and inclusion of disadvantaged groups. European guidance and related literature therefore frame social procurement as a legitimate component of strategic public procurement, while also stressing that implementation remains difficult and that price-centred routines still dominate in many practical settings. In Slovakia, this debate takes a more specific institutional form through Registered Social Enterprises (RSEs). Registered social enterprises are regulated by the Social Economy and Social Enterprises Act, while their procurement participation is governed by the Public Procurement Act (Zákon č. 112/2018 Z. z., 2018; Zákon č. 343/2015 Z. z., 2015).

The present article approaches the topic from a descriptive data-analytic perspective. Using big data from the Slovak Electronic Contracting System (EKS), it examines how procurement outcomes differ between contracts with and without RSE participation and how these outcomes vary within the RSE segment itself. The article concentrates on heterogeneity visible in observed data, especially in relation to savings, competition intensity, estimated contract size and CPV structure.

2. Literature review

Social enterprises are usually understood as organizations that combine economic activity with an explicit social mission, often linked to employment creation, inclusion, or community benefit. In the European discussion, this has been increasingly connected to public procurement, because procurement is no longer seen only as a transactional process focused on the lowest price, but also as a potential policy instrument for achieving broader social outcomes. As OECD/EU (2017) and the European Commission (2021) suggest, the involvement of social enterprises in procurement may support social inclusion, regional development, and socially responsible purchasing. In the Slovak setting, this discussion takes a more specific institutional form through *Registered Social Enterprises* (RSEs), whose status is linked to measurable social impact and to a legal framework that enables reserved contracts and other supportive procurement mechanisms. (Zákon č. 112/2018 Z. z., 2018)

The literature also shows that social enterprises face persistent barriers in procurement markets, even where the policy framework is formally supportive. Mazzei et al. (2017) and Steiner et al. (2019) emphasize that contract sizes are often too large for smaller social enterprises and that procurement structures may favour larger and better-capitalized organizations. Similar concerns are raised by Licite-Kurbe et al. (2021), who point to limited procurement readiness, insufficient professionalization, and low awareness among public authorities as important constraints. At the same time, the literature does not portray social enterprises only as weak or marginal actors. Sarter (2024), Buzzetto et al. (2020), and Fuentes-Bargues et al. (2021) suggest that their distinctive role may be especially relevant where procurement is linked to employment of disadvantaged groups or to broader social value objectives. The literature suggests a tension in that social enterprises may align well with the goals of social procurement, yet their actual access to markets still depends on procurement design, contract structure, and institutional capacity.

The Slovak evidence on this topic remains limited, but article by Delina and Šulová (2025) provides an important starting point. Using the same broader data environment, the article showed that RSE participation in Slovak public procurement is still low, that successful outcomes are concentrated among a relatively small number of suppliers, and that activity is clustered in selected CPV categories. At the same time, Delina and Šulová (2025) framed these findings mainly in terms of untapped potential and the possible use of AI-driven matching to improve supplier visibility and access. The present article builds on that empirical baseline, but moves in a different direction.

A further reason for focusing on heterogeneity comes from the wider procurement-outcomes literature. A consistent finding is that the number of bidders is one of the most important observable indicators related to competition and procurement efficiency. García Lafuente et al. (2021) show that the effect of competition may increase up to a certain number of bidders and then stabilize, while Tatrai et al. (2024) caution that bidder counts alone do not capture the whole competitive environment because procedure design, lot division, negotiation possibilities, and award criteria also matter. The same literature suggests that contract size and procurement categories should not be treated as secondary variables. Delina et al. (2021) and Drake et al. (2025) indicate that larger contracts may attract more bidders, but may not necessarily yield proportionally higher savings. This is important for the present study because observed differences between RSE and non-RSE contracts may partly

reflect differences in competition intensity, contract size, and category composition rather than a single uniform “RSE effect.”

3. Descriptive analysis of RSE behaviour and procurement outcomes

This section examines how procurement environments differ between contracts with RSE participation and non-RSE contracts, and then explores the internal heterogeneity of RSE outcomes across product categories and firms. In contrast to earlier work (Delina and Šulová, 2025) that primarily documented the extent and concentration of registered social enterprise (RSE) participation in public procurement, the present analysis places greater emphasis on observed procurement characteristics, including contract values, competition intensity, savings and firm level category specialization.

The dataset consists of more than 86,000 public procurement contracts from the Slovak Electronic Contracting System (EKS) over the period 2019–2025, covering almost all of the available data on EKS contracts for the period, excluding only contracts with missing or incorrectly entered values. Of the whole dataset, only 1770 contracts (2%) involve RSE participation.

3.1. Competition, savings, and contract size

The descriptive comparison indicates that contracts with RSE participation differ from non-RSE contracts in several notable respects. First, the mean number of bidders is substantially higher in contracts with RSE participation than in non-RSE contracts (approximately 5.5 versus 2.9), pointing to a more competitive procurement environment. Second, observed savings are also higher on average in contracts with RSE participation (approximately 26%) than in non-RSE contracts (approximately 14%). By contrast, the mean estimated contract value is remarkably similar in the two groups, amounting to roughly EUR 17.6 thousand for non-RSE contracts and EUR 17.4 thousand for contracts with RSE participation. Taken together, these descriptive results suggest that the presence of RSEs is not associated with systematically smaller contracts, but rather with contracts that appear, on average, to attract more bidders and exhibit higher price reductions relative to estimated value.

At the same time, these aggregate contrasts should be interpreted carefully. Higher competition and higher savings in contracts with RSE participation do not in themselves reveal the underlying mechanism. They may reflect differences in market segments, contract composition, product standardization, regional structure, procurement design, or bidder strategies. For that reason, the aggregate comparison is best understood as a starting point that motivates closer examination of heterogeneity within the RSE segment.

The boxplots in Figure 1 reinforce the basic descriptive picture. The distribution of the number of bidders shows that contracts with RSE participation are shifted upward relative to non-RSE contracts, not only in mean terms but also across a substantial part of the distribution. The same applies to savings: contracts with RSE participation tend to display higher typical savings, although both groups contain considerable dispersion and a visible upper tail of more extreme outcomes. In contrast, the figure on estimated values suggests that contract size is broadly comparable across the two groups, with no strong descriptive indication that RSE participation is limited only to very small contracts.

From the perspective of procurement outcomes, this combination is noteworthy. If contract sizes seems to be similar on average based on descriptive analysis, but contracts with RSE participation exhibit both higher competition and higher savings, then the difference is less likely to be driven by simple scale effects alone. At the same time, the spread within the RSE group is wide enough to suggest that the segment is internally diverse. Some RSE-related contracts appear in highly

competitive environments with pronounced savings, whereas others resemble more standard procurement patterns. This internal diversity becomes clearer once the analysis moves from the aggregate level to individual product categories and firms.

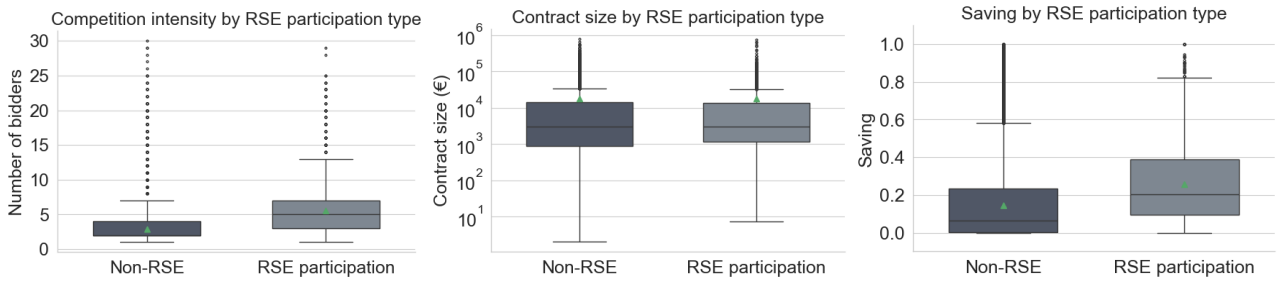


Figure 1. Competition, contract size and saving comparison between contracts involving RSE firms and contracts without RSE participation Source: authors

3.2. Heterogeneity of outcomes across CPV categories

The next step is to examine whether the observed RSE profile is uniform across procurement categories. The distribution of savings across CPV categories among successful RSE contracts indicates substantial heterogeneity, as seen in Figure 2. Some categories are associated with relatively high median savings and wider dispersion, whereas others show more modest savings and tighter distributions. The number of observations also varies considerably across categories, which cautions against overinterpreting any single category in isolation. Nevertheless, the graph suggests that RSE outcomes are not homogeneous across the product spectrum and may depend on the type of goods or services procured.

This point is important because the average RSE difference observed in the previous subsection may mask very different procurement environments. In categories characterized by more standardized products, price-based competition may dominate and generate stronger savings. In other categories, where contract specification, customization, or service intensity plays a greater role, savings patterns may differ. At the descriptive level, the figure therefore indicate that CPV composition is a relevant source of heterogeneity and should be taken into account in any deeper analysis.

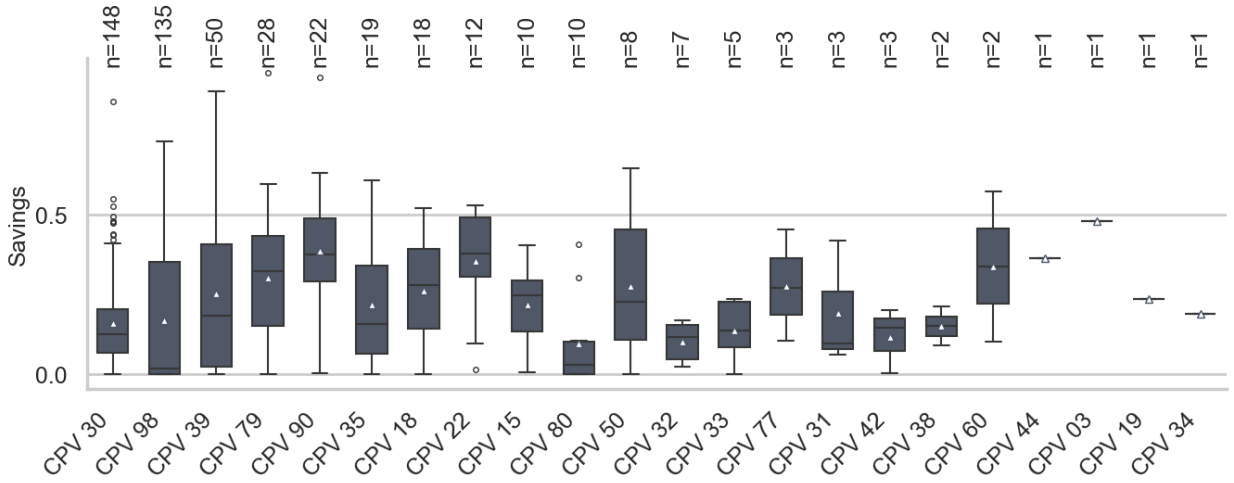


Figure 2. Distribution of savings across RSE winning CPV categories Source: authors

The category structure also points to concentration in particular segments. As shown by the leading CPV distributions, RSE activity is not evenly spread across procurement categories.

3.3. Firm-level heterogeneity and concentration among leading RSE contractors

Based on the descriptive evidence from the earlier analysis on similar dataset, there seems to be a strong concentration of few leading contractors among the contracts on the platform (Delina and Šulová, 2025). Procurement success is concentrated among a relatively small number of RSE firms, yet these leading firms do not appear to behave similarly. Adding to this information, the data also suggest that even on a contractor level, procurement categories might differ.

Table 1 shows substantial variation in the number of distinct CPV categories served by the top firms. At one end, *DC Sielnica* appears markedly diversified, with activity spanning a broad set of categories (12). *CNCteam* and *Registrovaný sociálny podnik Alfa* also display a wider category spread than several other firms. At the other end, firms such as *Autoškola Beňuš* and *ELSTROTE* are active in only one category in the observed period, while others such as *AMG Security*, *CHANGE COMPUTER*, *millhill*, and *WASCO – družstvo* remain comparatively focused. Although procurement outcomes are concentrated among top RSE contractors, the successful firms differ in the breadth of their market engagement.

Table 1. Unique CPV categories per firm (top 10 firms with highest contracts count)

Firm name	Different CPV categories (2nd level)	Most Frequent CPV	Total number of contracts
WASCO - družstvo	3	98	129
DC Sielnica, n. o., „r. s. p.“	12	30	104
CHANGE COMPUTER s. r. o.	3	30	50
Breco s. r. o.	4	39	36
CNCteam, s.r.o.	7	39	31
Registrovaný sociálny podnik Alfa s.r.o.	5	30	20
ELSTROTE, spol. s r.o.	1	18	10
AMG Security s.r.o.	2	79	9
millhill s.r.o.	3	39	8
Autoškola Beňuš, s. r. o.	1	80	8

Source: authors

This heterogeneity is relevant for interpretation of aggregate results. If a large share of RSE procurement activity is carried by firms with different category portfolios, operating modes, and observed savings profiles, then average statistics for the whole RSE group are unlikely to describe a single typical procurement pattern. Rather, the group seems to contain at least several distinct firm-level strategies or market positions.

At the same time, if we look at the most frequent CPV categories among the top 10 RSE suppliers, despite the overall diversity it seems that social enterprises often compete in highly standardized markets (*CPV 30 - Office and computing machinery, equipment and supplies except furniture and software packages* and *CPV 39 - Furniture (incl. office furniture), furnishings, domestic appliances (excluding lighting) and cleaning products*). This is in contrast with the popular belief that RSE firms prefer to operate in CPV categories that are highly specialized. Standardized procurement might also naturally attract a wider pool of commercial bidders which could provide, at least in part, a logical explanation to the higher average number of bidders in contracts where RSE firms participate. However, testing such hypothesis would require a further analysis.

3.4. Brief illustrative firm profiles

A few short firm-level examples of the top 3 winning RSE firms, based on contract counts on the EKS platform, help to clarify why a more detailed follow-up analysis may be warranted.

DC Sielnica is notable for its high number of contracts (more than 100) and broad category spread. The observed portfolio includes a wide variety of products and services, such as computers, printers, construction works, airsoft equipment, extending beyond a narrowly defined core procurement niche. This is especially interesting considering the firm's history being a nursery and elementary school. At the same time, its typical savings appear lower than the average observed in the broader RSE-participation group. Descriptively, this makes the firm an interesting case because it combines high activity with broad scope and comparatively moderate savings.

CHANGE COMPUTER also records a relatively high number of contracts (50), but its profile appears more concentrated in information technology-related procurement. In descriptive terms, it resembles an e-commerce-oriented supplier operating in a standardized segment where price comparison may be relatively straightforward. Yet, similar to the previous case, its typical savings do not appear especially high compared with the broader RSE-participation distribution. This does not imply any irregularity in itself, but it indicates that repeated success may emerge under procurement conditions that differ from the average RSE contract.

WASCO – družstvo is the top contractor from the contracts count perspective (129), however it stands out for a different reason. Its profile appears more closely aligned with what might be expected from a social-enterprise-oriented niche, especially through concentration in CPV category 98 - *Other community, social & personal services*. At the same time, its activity is more often observed in less contested environments, and the associated savings tend to be on the lower side. In the present descriptive framework, this profile is interesting precisely because it differs from the broader pattern of higher competition and higher savings seen in the aggregate RSE-participation group.

4. Directions for further research

The descriptive results show that contracts with RSE participation differ from non-RSE contracts on average, but they also show that the RSE segment is internally heterogeneous across categories and firms. For that reason, two further research directions appear interesting.

4.1. Causal analysis in heterogeneous procurement environments

A first direction is to move from descriptive comparison toward causal analysis of procurement outcomes in heterogeneous environments. The central question would not simply be whether RSE participation is associated with higher savings or stronger competition, but under what conditions such differences arise and how they vary across contract types, categories, and bidder structures.

Work on heterogeneous treatment effects offers a useful methodological starting point here. Wager and Athey (2018) for example propose advanced methods of using machine learning based causal forests for estimation of the heterogenous treatment effects. Such approaches make it possible to move beyond simple correlations and identify how the effect of RSE participation varies across different procurement settings while reducing bias from observed confounding factors. This can provide a more credible explanation of the mechanisms through which RSE participation influences procurement outcomes and help us, to some extent, distinguish genuine treatment effects from differences driven by underlying contract or market characteristics.

More broadly, these developments align with the growing use of data-driven and artificial intelligence methods in supply chain management, where understanding heterogeneous responses across suppliers, contracts, and market conditions is becoming increasingly important for informed decision-making.

4.2. Anomaly-oriented analysis of firm behaviour

A second direction concerns the analysis of atypical or anomalous firm behaviour. Since the descriptive evidence suggests that successful RSE firms do not follow a single uniform pattern, it may be useful to identify a broadly typical RSE procurement profile and the firms that deviate markedly from it. For example, *DC Sielnica* can be considered one such illustrative case, given its unusually high contracts count, broad category dispersion, and comparatively moderate savings relative to the RSE group.

Future research could therefore employ anomaly-detection or clustering techniques to identify firms whose procurement behaviour differs substantially from the prevailing RSE pattern. Such deviations may reflect alternative business strategies, sector-specific conditions, or potentially possible misuse of advantages associated with social enterprise status without a corresponding level of social or economic performance.

The anomaly-detection literature provides a relevant methodological foundation for this purpose, including statistical and machine learning approaches designed to identify deviations from dominant patterns. Recent studies in public procurement demonstrate the potential of these methods for detecting irregular behaviour. For example, Fazekas et al. (2026) developed a Random Forest-based model combining multiple behavioural indicators, such as bidding patterns, pricing anomalies, and relational characteristics, achieving detection accuracies of 70–84% across seven European countries. Similarly, García Rodríguez et al. (2022) showed that machine-learning algorithms enhanced with statistical screening variables can identify anomalous bidding patterns with detection rates of 81–95% across multiple procurement datasets. These findings suggest that similar approaches could be adapted to identify RSE firms whose procurement profiles deviate substantially from typical patterns and to examine the underlying drivers of such deviations.

5. Conclusion

Overall, the descriptive evidence suggests that contracts with RSE participation are characterized by higher average competition and higher observed savings than non-RSE contracts, while average contract size remains broadly similar. At the same time, the RSE segment is not internally uniform. Meaningful differences are visible across product categories and among the leading firms, both in the breadth of procurement activity and in the structure of outcomes. The brief firm-level examples further indicate that some mechanisms may be firm-specific and deserve closer scrutiny. For that reason, the most appropriate reading of the present results is as a structured descriptive basis for more targeted causal and anomaly-oriented research, with potential relevance for AI-driven supply chain analytics.

References

Buzzetto, R. R., Bauli, M. R., & Carvalho, M. M. de. (2020). The key aspects of procurement in project management: investigating the effects of selection criteria, supplier integration and dynamics of acquisitions. Production, 30. https://doi.org/10.1590/0103-6513.20190112

Delina, R., Gróf, M., & Dráb, R. (2021). Understanding the Determinants and Specifics of Pre-Commercial Procurement. Journal of Theoretical and Applied Electronic Commerce Research, 16(2), 80-100. https://doi.org/10.4067/S0718-18762021000200106

Delina, R., & Šulová, M. (2025). The potential of social enterprises in Slovak public procurement: Descriptive analysis and AI-driven matching framework. In IDIMT-2025, ICT in Business: AI Everywhere? Glory and Disgrace of AI (pp. 327-336). https://dialnet.unirioja.es/servlet/articulo?codigo=10441444

Drake, S., & Xu, F. (2025). Contractual Requirements and Bidding Behavior in Public Procurement With Entry. Journal of Public Economic Theory, 27(1), e70019. https://doi.org/10.1111/jpet.70019

European Commission. (2021). Buying social: A guide to taking account of social considerations in public procurement (2nd ed.). Publications Office of the European Union. https://op.europa.eu/en/publication-detail/-/publication/47c69b3a-cfcf-11eb-ac72-01aa75ed71a1

Fazekas, M., Tóth, B., Wachs, J., & Abdou, A. (2026). Public procurement cartels: A large-sample testing of screens using machine learning. International Journal of Industrial Organization, 104, 103228. https://doi.org/10.1016/j.ijindorg.2025.103228

Fuentes-Bargues, J. L., Bastante-Ceca, M. J., Ferrer-Gisbert, P. S., & González-Cruz, M. C. (2021). Analysis of the Situation of Social Public Procurement of Works at the Valencian Region (Spain). Sustainability, 13(1), 175. https://doi.org/10.3390/su13010175

García Lafuente, L. M., & Mochón Sáez, A. (2022). Competition effects in EU external aid supply tenders funded with the Pre-accession and Neighbourhood instruments. Empirica, 49(2), 461-484. https://doi.org/10.1007/s10663-021-09518-5

García Rodríguez, M. J., Rodríguez-Montequín, V., Ballesteros-Pérez, P., Love, P. E. D., & Signor, R. (2022). Collusion detection in public procurement auctions with machine learning algorithms. Automation in Construction, 133, 104047. https://doi.org/10.1016/j.autcon.2021.104047

Licite-Kurbe, L., & Gintere, D. (2021). Analysis of Financial Support Instruments for Social Enterprises in Latvia. Rural Sustainability Research, 45(340), 76-84. https://doi.org/10.2478/plua-2021-0009

Mazzei, M., & Roy, M. J. (2017). From Policy to Practice: Exploring Practitioners' Perspectives on Social Enterprise Policy Claims. VOLUNTAS: International Journal of Voluntary and Nonprofit Organizations, 28(6), 2449-2468. https://doi.org/10.1007/s11266-017-9856-y

OECD, & EU (2017). Boosting Social Enterprise Development: Good Practice Compendium. Local Economic and Employment Development, 2016. https://doi.org/10.1787/9789264268500-en

Sarter, E. K. (2024). Public procurement and labour market inequality: Conceptualising a multi-faceted relationship. Journal of Industrial Relations, 66(1), 130-152. https://doi.org/10.1177/00221856231221417

Steiner, A., & Teasdale, S. (2019). Unlocking the potential of rural social enterprise. Journal of Rural Studies, 70, 144-154. https://doi.org/10.1016/j.jrurstud.2017.12.021

Tátrai, T., Vörösmarty, G., & Juhász, P. (2024). Intensifying Competition in Public Procurement. Public Organization Review, 24(1), 237–257. https://doi.org/10.1007/s11115-023-00742-0

Wager, S., & Athey, S. (2018). Estimation and inference of heterogeneous treatment effects using random forests. Journal of the American Statistical Association, 113(523), 1228-1242. https://doi.org/10.1080/01621459.2017.1319839

Zákon č. 112/2018 Z. z. o sociálnej ekonomike a sociálnych podnikoch a o zmene a doplnení niektorých zákonov, Zbierka zákonov Slovenskej republiky (2018). https://www.slov-lex.sk/ezbierky/pravne-predpisy/SK/ZZ/2018/112/

Zákon č. 343/2015 Z. z. o verejnom obstarávaní a o zmene a doplnení niektorých zákonov, Zbierka zákonov Slovenskej republiky (2015). https://www.slov-lex.sk/ezbierky/pravne-predpisy/SK/ZZ/2015/343/

GENERATIONAL ASYMMETRY IN AI ADD-ON LAYER IMPLEMENTATION FOR ERP: A LONGITUDINAL CASE STUDY IN A SPECIFIC MANUFACTURING ENTERPRISE

Vladislav Kašpar

Pan - European University
xkaspar.vladisl@is.peuni.cz

DOI: 10.35011/IDIMT-2026-307

Keywords

Artificial Intelligence, ERP, Baby Boomers, Generation X, Y and Z, technology adoption.

Abstract

The emergence of Generations Y and Z into decision-making positions creates growing pressure within companies for digitalization and flexibility of business processes – a challenge that existing ERP systems, designed primarily for the communication and work patterns of older generations, are unable to address quickly or cost-effectively. This study examines how the implementation of artificial intelligence (AI) tools as an add-on layer on top of ERP systems can serve as a response to generationally conditioned diversity in process and communication expectations, and identifies the key risks associated with this implementation.

1. Introduction

Digital transformation of business processes represents one of the key strategic challenges facing the contemporary manufacturing sector. Companies face growing pressure to increase process efficiency and adaptability, and the rise of artificial intelligence (AI) tools opens new possibilities for meeting these demands (Davenport & Ronanki, 2018; Vial, 2019). In parallel with this development, organizations are undergoing a significant demographic transformation: Generations Y and Z are occupying decision-making positions and bringing different expectations regarding communication channels and process management methods (Cennamo & Gardner, 2008; Twenge, 2017).

2. Theoretical framework

The theoretical section rests on two pillars that define the framework for analysing generational differences in the integration of AI into ERP systems. ERP systems constitute the integration backbone of business processes; however, their architecture is typically associated with a high degree of process standardization, modification complexity, and the risk of organizational-technological misalignment (Amid et al., 2012; Klaus et al., 2000; Soh & Sia, 2004). Intergenerational perceptions of working approaches differ, and it is therefore necessary to identify common acceptable elements.

2.1. Pillar P1: Definition of generational cohorts and their work values

Generational theory is based on the assumption that individuals born in the same historical period share formative experiences that shape their value systems, work attitudes, and preferences towards organizational processes (Cennamo & Gardner, 2008). In the context of today's labour market, four generational cohorts are typically distinguished: Baby Boomers (born approx. 1946–1964), Generation X (born approx. 1965–1980), Generation Y (born approx. 1981–1996), and Generation Z (born after 1997). In the context of ERP systems, these generational differences manifest in specific ways: Baby Boomers and Gen X interact with ERP primarily through structured forms, workflows, and modules whose logic they helped shape over the years. Generations Y and Z, by contrast, prefer conversational and visual interfaces, seek direct ERP integration with external tools (chat, mobile applications), and expect immediate feedback. The implementation of AI as an ERP add-on layer is therefore ambiguous from the perspective of generational diversity: it may increase system accessibility for younger generations while simultaneously disrupting the established process routines of older cohorts. Literature on the impact of automation on the labour market documents that the introduction of AI and robotics triggers a specific type of job insecurity among older workers that goes beyond general technophobia (Alcover et al., 2021).

2.2. Pillar P2: AI in business processes - Opportunities and systemic risks

One of the key risks of AI deployment in business processes - yet one frequently underestimated in implementation practice - is the gradual erosion of natural communication patterns through which tacit knowledge is created and shared within organizations. Nonaka and Takeuchi (1995) demonstrated in the SECI model that the most valuable component of organizational knowledge - tacit knowledge - arises primarily through the process of socialization: direct interaction between individuals and the sharing of work situations. It is precisely this process that is structurally threatened in AI-mediated environments: if an employee uses an AI output as a substitute source of contextual information rather than consulting a colleague, no social exchange takes place through which shared understanding would form and new ideas would be generated. It must be acknowledged that the erosion of tacit knowledge in organizations is not caused exclusively by the advent of AI.

3. Methodology

Overarching research question

How does generationally conditioned asymmetry in the perception of AI threats and benefits influence the course and outcomes of implementing AI as an add-on layer to an ERP system in a manufacturing enterprise?

This question is deliberately formulated as a how-question rather than a whether-question, and corresponds to the qualitative case study design whose aim is to understand mechanisms rather than to test statistical relationships (Eisenhardt & Graebner, 2007; Yin, 2018). The focus on the manufacturing sector is deliberate: manufacturing enterprises are characterized by high process complexity, deep integration of ERP into operational management, and a typically multigenerational workforce composition (concurrent presence of all four cohorts). This combination creates ideal conditions for studying generationally conditioned asymmetry in AI adoption.

Sub-Research questions

- RQ1: How do employees from different generational cohorts perceive the benefits and threats of an AI add-on to ERP?

- RQ2: How do these perceptions affect the willingness to share process know-how and participate in AI implementation?
- RQ3: What impacts does the AI add-on to ERP have on process efficiency, communication, and the creation of tacit knowledge?

3.1. Research design

The research is conceived as a comparative longitudinal single-case study of a specific manufacturing enterprise operating in the field of custom production. Data were collected in two periods: in April 2024 at the outset of the AI add-on ERP implementation, and in April 2025, after one year of more advanced operational use.

3.2. Case selection

Case selection was purposive (purposive sampling; Patton, 2015) and governed by pre-established criteria of relevance, accessibility, and comparability. Selection was guided by three criteria: (a) selection of key employees directly involved in defining and managing AI implementation, (b) representation of all four generational cohorts present in the enterprise, and (c) daily interaction with both systems under study (ERP and the AI add-on). Nine informants is consistent with the standards of qualitative case study research, where the relevance of the sample takes precedence over its size. The aim was analytical saturation of key themes. Given that the enterprise under study has a limited number of employees directly involved in integrating AI into ERP, the nine informants constitute a relevant sample. Their breakdown by generational cohort and job position is described in Table 1.

3.3. Data collection methods

Data were collected through semi-structured interviews and informal discussions within the research questions framework. The interview protocol was developed on the basis of the conceptual model and covered thematic areas relating to the three research questions: RQ1–RQ3. Nine interviews were conducted with informants representing three levels according to the inclusion criteria. Interviews lasted 30–45 minutes and were reviewed and authorized by informants after transcription. The first-phase data collection (T0) took place during April 2024, immediately after the AI add-on was launched into live operation. The second phase (T+12) took place in April 2025 with the same nine informants, enabling direct longitudinal comparison. The identical sample composition across both phases is a deliberate methodological choice ensuring intersubjective data consistency (Yin, 2018). The interviews were not analysed as a quantified data set; their content was recorded in the form of structured notes, which were subsequently authorized by the informants. The findings are not presented as statistically generalizable conclusions, but as an analytical interpretation.

3.4. Analytical strategy

The study employs qualitative methods because the research design seeks to understand mechanisms and processes rather than to measure frequencies. Thematic analysis (Braun & Clarke, 2006) was used as the specific analytical method: identification of themes, and their interpretation in the context of the research questions. Quantitative statistical methods were not applied, as they are inadequate for the qualitative design chosen and the nature of the research questions.

3.5. Validity and reliability

The quality of the research design was assessed along two dimensions: internal validity and reliability. Internal validity - causal inferences were formulated as analytically grounded interpretations. Reliability - the repeatability of the research process - was ensured through a protocol containing the interview guide, observer instructions, and a document analysis procedure.

4. Results

4.1. Research sample profile

The research sample consists of nine informants purposively selected on the basis a generational cross-section encompassing all four cohorts present in the enterprise, and the intensity of daily interaction with both the AI system and ERP. An overview of the sample is provided in Table 1.

Table 1. Research sample profile

Code	Generation	Role	Primary area of AI use	Research phase
BB-1	Baby Boomers	Project manager	Technological calculations and order entry into ERP	T0, T+12
BB-2	Baby Boomers	Operations manager	Analytical calculations, manufacturing equipment performance	T0, T+12
X-1	Gen X	Quality manager	Documentation creation and processing of customer documentation	T0, T+12
X-2	Gen X	Finance manager	Digitalization and automation of document receipt	T0, T+12
X-3	Gen X	Enterprise director	Process data analytics, strategic evaluation	T0, T+12
Y-1	Gen Y	Sales representative	Order retrieval, creation of sales offers in ERP	T0, T+12
Y-2	Gen Y	IT manager	Programming and implementation of ERP functionality	T0, T+12
Z-1	Gen Z	Sales representative	Comprehensive sales agenda and customer communication	T0, T+12
Z-2	Gen Z	Developer/programmer	Design and automation of ERP subcomponents	T0, T+12

Source: author

4.2. RQ1: Generational diversity in process expectations and AI adoption

4.2.1. Approach to AI outputs: Verifiability vs. flexibility

A markedly different relationship to AI outputs is evident across generational groups. Informants from Generations BB and X consistently emphasize - in both phases - the need for verifiability and control. BB-1 describes his role vis-à-vis the AI system in both phases as one of checking and correcting: he accepts the output only after verification.

Conversely, informants from Generation Z exhibit a markedly more flexible approach. Z-1 entrusts her entire sales agenda to AI and does not evaluate the outputs declaratively. Z-2, in T+12, describes

an active iterative approach - entering the anticipated result and tuning the path to achieve it. Y-1, in T+12, declares complete trust in the outputs.

4.2.2. Perception of intergenerational differences in AI use

Seven of the nine informants explicitly perceive that younger colleagues use the AI system differently or more intensively. This perception differs qualitatively according to the generational membership of the evaluating informant. BB informants critically evaluate the greater flexibility of Y/Z as a lack of rigour and uncritical acceptance of outputs.

Noteworthy are the statements of X-2 and X-3, who challenge the generational schema. X-2, in both phases, states that age is not decisive and that mental attitude is the determining factor. X-3, however, revises his view in T+12: "It is more about the level of use and the mental attitude of individual colleagues regardless of age."

This shift in interpretation suggests that with growing experience with AI, individual variability in adoption becomes analytically more relevant than generational membership as such - a finding consistent with the critique of strict generational categorization in the management literature (Twenge, 2017).

4.2.3. Process ambitions: What respondents expect from the system

Responses to the question of what the system would need to be capable of for more intensive use reveal a fundamental polarization of process ambitions between the BB+X and Y+Z groups. BB and X informants formulate requirements oriented towards reliability and process consistency: error-free operation, verifiability of calculation methodology against ERP databases for validation. Y and Z informants, by contrast, demand expansion of the system's functional capacity: solving more complex tasks and autonomous handling of communication.

This polarization within the case under study suggests that a system designed primarily for one group will be structurally underperforming for the other, thereby further deepening the generational imbalance in adoption.

4.3. RQ2: Threat perception and tacit know-how sharing

The second sub-research question focuses on the relationship between generationally conditioned threat perception from AI implementation and the willingness to share tacit process know-how with the AI system. This dimension yields the analytically most significant findings of the entire study.

4.3.1. Perception of existential threat: Generational asymmetry and longitudinal shift

In T0, both BB informants and two of the three X informants (X-1, X-2) perceive potential displacement by AI as real or probable. X-3 and Y-2 reject any threat; Z-1 adopts a pragmatic stance referencing the appropriation of AI outputs as one's own work.

In T+12, shifts occur. BB-1 points to the quality of his own work as an irreplaceable advantage over AI, and X-1 and X-2 reduce or specify the degree of threat in T+12 - instead of direct job displacement, they identify the potential for headcount reduction within the department.

4.3.2. Tacit know-how sharing: The key research finding

The most powerful and analytically most significant finding of the study is a marked and stable generational bifurcation in the willingness to share tacit process know-how with the AI system.

The pattern appears consistent: all four Y+Z informants share know-how in both phases; four of the five BB+X informants do not share (the exception being X-3 in T+12). This pattern provides empirical support for the knowledge hoarding mechanism assumed by the conceptual model and described in the literature (Cabrera & Cabrera, 2002; Connelly & Kelloway, 2003).

4.3.3. Trust in AI outputs: A reliability gradient

The degree of trust in AI outputs mirrors the know-how sharing pattern. BB-1 declares distrust in both phases; Gen X exhibits mixed attitudes with a slight increase in trust. In T+12, X-1 transitions from selective verification to full trust. Generations Y and Z display a higher degree of trust, which increases further in T+12. Among Y/Z informants, signals of automation bias are evident (Parasuraman & Manzey, 2010).

4.4. RQ3: Opportunities and risks of AI implementation as an ERP add-on

The third sub-research question (RQ3) maps the specific opportunities and risks that informants associate with AI implementation. The longitudinal comparison reveals a marked shift in the overall assessment of the implementation and in the breadth of identified benefits.

4.4.1. Identified opportunities

Data from T0 show selective and generationally asymmetric perception of benefits. Y+Z informants and Y-2 are already able, in T0, to name specific benefits: faster ERP data entry, customer communication, programming assistance. BB-1, in T0, explicitly evaluates AI as a factor that complicates work; X-2 does not yet perceive any benefit in T0.

In T+12, the spectrum of benefits is considerably broader and includes respondents from BB. BB-1, who identified no benefit in T0, describes in T+12 concrete values - rapid processing of bulk calculations and technological problem-solving. BB-2 appreciates AI's analytical and comparative capabilities for monitoring manufacturing equipment performance. X-2 identifies a reduction in the volume of routine tasks.

4.4.2. Identified risks

Across both phases, three categories of risk are repeatedly identified.

The first risk is sensitivity to input quality, including hallucinations. Informants repeatedly mention that AI requires precise and structured input; with non-standard tasks or when using open data sources, it produces erroneous outputs. Z-2, in T+12, explicitly identifies hallucinations as an operational risk.

The second risk is erosion of independent judgment and automation bias. X-1, in T+12, states that her supervisor requires her own analytical perspective because AI output is recognizable. Z-1, in both phases, declares that she accepts AI outputs uncritically and presents them as her own results.

The third risk is organizational resistance as a blocker of implementation. X-3, in T0, identifies collegial conflict as a key operational problem.

4.4.3. Customer and communication dimension: An empirically grounded risk

A specific risk repeatedly signalled by the data is the degradation of customer communication. Z-1, in both phases, mentions customer complaints about confused or unnatural communication mediated by AI. This statement is analytically relevant to the commercial reach of implementation - one of the

two outcome dimensions of the conceptual model - and suggests that the automation of customer communication brings efficiency benefits while simultaneously posing a risk of loss of authenticity and relational quality of interaction.

5. Discussion

The findings from RQ1 broadly confirm the predictions of the UTAUT2 model (Venkatesh et al., 2012): informants from Generations Y and Z exhibit motivation and technological behaviour that supports rapid and comprehensive adoption of the AI system; informants from Generations BB and X display a consistent need for verifiability and transparency of outputs.

The polarization of process ambitions - BB+X demanding reliability of existing processes, Y+Z demanding functional expansion - is a finding that can be derived from two of the model's dimensions: performance expectancy and hedonic motivation.

The most robust finding of the study is the generational bifurcation in the willingness to share tacit know-how with the AI system: 4/4 Y+Z informants share, 4/5 BB+X informants do not share, and this pattern is stable across both time points. This provides support - within the case under study - for the central prediction of the conceptual model, as well as for the knowledge hoarding mechanism described by Cabrera and Cabrera (2002) and Connelly and Kelloway (2003).

The author identifies the bidirectional feedback loop as the principal theoretical contribution. The original conceptual model assumed a unidirectionally driven feedback loop: BB+X generations withhold tacit know-how → AI model is trained on incomplete data → output quality is lower → BB+X distrust is confirmed → resistance deepens. This mechanism is supported by the study's empirical data. The study, however, identifies a complementary active mechanism that the conceptual model did not explicitly anticipate automation bias in Generation Y+Z, who in both phases declare that they accept AI outputs uncritically and present them as their own work. In both T0 and T+12, customers report complaints about confused communication mediated by AI. This chain - uncritical consumption of AI outputs → externalization of erroneous or impersonal outputs → negative feedback from customers or supervisors → legitimization of BB+X distrust - constitutes the active component of the feedback loop.

Three key recommendations follow from the findings described above. (1) Implementation strategy must be generationally differentiated: for BB+X, ensure output verifiability; for Y+Z, build in structural verification steps. (2) Tacit know-how transfer requires explicit mechanisms such as validation dialogues or know-how review sessions, linked to incentive structures. (3) The communication framework should present AI as a tool that enhances the professional relevance of experienced workers, rather than as a threat to their positions (Alcover et al., 2021; Cabrera & Cabrera, 2002).

6. Conclusion

This study set out to examine how the implementation of AI tools as an add-on layer to an ERP system enables a manufacturing enterprise to respond to intergenerationally diversified process and communication expectations, and to identify the key opportunities and risks associated with this implementation. From a theoretical standpoint, the most original contribution is the identification of a bidirectional feedback loop as a mechanism of sub-optimization of AI implementation in a generationally diverse environment. The study extends the original conceptual model with an active

component: automation bias in Y+Z produces erroneous outputs that legitimize BB+X distrust, whereby the feedback loop acquires an asymmetrically combined character.

References

Alcover, C.-M., Guglielmi, D., Depolo, M., & Mazzetti, G. (2021). Aging-and-tech job vulnerability. Organizational Psychology Review, 11(2), 175–201. https://doi.org/10.1177/2041386621992105

Amid, A., Moalagh, M., & Zare Ravasan, A. (2012). Identification and classification of ERP critical failure factors in Iranian industries. Information Systems, 37(3), 227–237.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative Research in Psychology, 3(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Cabrera, A., & Cabrera, E. F. (2002). Knowledge-sharing dilemmas. Organization Studies, 23(5), 687–710.

Cennamo, L., & Gardner, D. (2008). Generational differences in work values. Journal of Managerial Psychology, 23(8), 891–906.

Connelly, C. E., & Kelloway, E. K. (2003). Predictors of employees' perceptions of knowledge sharing cultures. Leadership & Organization Development Journal, 24(5), 294–301.

Davenport, T. H., & Ronanki, R. (2018). Artificial intelligence for the real world. Harvard Business Review, 96(1), 108–116.

Eisenhardt, K. M., & Graebner, M. E. (2007). Theory building from cases. Academy of Management Journal, 50(1), 25–32. https://doi.org/10.5465/amj.2007.24160888

Klaus, H., Rosemann, M., & Gable, G. G. (2000). What is ERP? Information Systems Frontiers, 2(2), 141–162.

Nonaka, I., & Takeuchi, H. (1995). The knowledge-creating company. Oxford University Press.

Parasuraman, R., & Manzey, D. H. (2010). Complacency and bias in human use of automation. Human Factors, 52(3), 381–410. https://doi.org/10.1177/0018720810371093

Patton, M. Q. (2015). Qualitative research and evaluation methods (4th ed.). SAGE Publications.

Soh, C., & Sia, S. K. (2004). An institutional perspective on sources of ERP package-organisation misalignments. Journal of Strategic Information Systems, 13(4), 375–397.

Twenge, J. M. (2017). iGen. Atria Books.

Venkatesh, V., Thong, J. Y. L., & Xu, X. (2012). Consumer acceptance and use of information technology: Extending the unified theory of acceptance and use of technology. MIS Quarterly, 36(1), 157–178.

Vial, G. (2019). Understanding digital transformation: A review and a research agenda. Journal of Strategic Information Systems, 28(2), 118–144. https://doi.org/10.1016/j.jsis.2019.01.003

Yin, R. K. (2018). Case study research and applications (6th ed.). SAGE Publications

COMPOSITION EFFECTS IN THE UTILITY GREENIUM

Daniel Kratky, Robert Verner

Bratislava University of Economics and Business, Slovakia
Faculty of Business Economics with seat in Kosice
daniel.kratky@euba.sk, robert.verner@euba.sk

DOI: 10.35011/IDIMT-2026-315

Keywords

ESG bonds; greenium; utility sector; primary market; credit spreads; sustainable finance.

Abstract

The utility sector is responsible for a large share of global CO2 emissions but is also the principal vehicle of the energy transition, and its issuers have become among the most active users of labelled debt. This study investigates whether ESG-labelled bonds, comprising green, social, sustainable and sustainability-linked structures issued by utility companies between January 2020 and April 2026, are priced at a conditional spread differential relative to otherwise comparable conventional bonds. Using a primary-market data set of 2,725 investment-grade and high-yield utility bond tranches covering about USD 1,635 billion of notional issuance, we estimate cross-sectional ordinary least squares models with heteroskedasticity-robust standard errors and a rich battery of controls and fixed effects. Unconditionally, ESG-labelled bonds price roughly 10 basis points inside conventional peers. Once we control for credit rating, tenor, issue size, benchmark, region, currency and year fixed effects, the sign reverses: ESG-labelled bonds are associated with launch spreads about 12 basis points wider than matched conventional peers. The association is concentrated in North American issuance and in the tightening cycle from 2022 onwards. The findings qualify the conventional narrative of a universal utility-sector greenium and suggest that composition effects explain most of the headline pricing advantage.

1. Introduction

The market for sustainable debt has grown from a niche product into a mainstream funding instrument in less than a decade. According to the Climate Bonds Initiative, cumulative labelled issuance passed the USD 5 trillion threshold in 2024, and utility issuers consistently rank among the three most active industries. The utility sector is at the centre of both the climate problem and its solution: electricity and gas producers account for roughly 40% of global energy-related emissions, while simultaneously financing the largest share of renewable-generation and grid-upgrade capital expenditure. This dual role makes utilities a particularly informative laboratory in which to test whether sustainable labelling translates into a measurable pricing advantage in the primary bond market. Beyond their direct climate role, utilities also occupy a structural position in the energy supply chain: their capital cost shapes the financing terms for transmission and distribution networks, for renewable-generation procurement, and for the transition-related supplier base. Variation in the cost of labelled debt for utilities therefore propagates through procurement decisions, the resilience of energy infrastructure,

and the bankability of long-dated supplier contracts that underpin the broader sustainable supply chain.

A large literature documents a so-called greenium, typically estimated at 1 to 9 basis points (bps) for investment-grade green bonds across sectors and currencies. Yet empirical estimates are far from uniform: they depend on sample composition, the treatment of issuer fixed effects, the choice of benchmark, and the macro-financial regime in which the bonds are priced. After the 2022 monetary-tightening cycle, anecdotal evidence from syndicate desks suggested that greeniums had compressed or even disappeared, as higher real yields brought conventional investors back to the market and weakened the relative scarcity value of labelled paper. This paper contributes to the greenium debate in four ways. First, we focus exclusively on the utility sector, which removes the sector heterogeneity that has complicated prior cross-sector studies. Second, our sample of 2,725 deals is, to our knowledge, among the largest single-sector samples assembled to date and spans a complete monetary cycle. Third, we distinguish between four ESG sub-labels, namely green, social, sustainable and sustainability-linked, to test whether the market prices these structures differently. Fourth, we position the findings within the literature on sustainable supply-chain finance and the cost of transition capital. The remainder of this paper is organised as follows. Section 2 reviews the relevant literature. Section 3 formulates the hypotheses. Section 4 describes the data and the empirical methodology. Section 5 presents the results. Section 6 discusses implications and limitations, and Section 7 concludes.

2. Literature review

The pricing of green and other labelled bonds is a comparatively young but rapidly expanding strand of the sustainable-finance literature. Early studies such as Zerbib (2019) and Baker et al. (2022) document a modest negative yield differential between green and matched conventional bonds in the range of 2 to 9 bps, interpreting this greenium as a non-pecuniary preference of investors for environmentally-labelled debt. Flammer (2021) links the effect to investor demand from dedicated sustainable funds and to a signalling channel that lowers information asymmetry between issuer and investor. More recent contributions qualify the magnitude and persistence of the greenium. Caramichael and Rapp (2024) find that the greenium has shrunk as the labelled-bond market has matured, while Pietsch and Salakhova (2022) show that the premium in the euro area is concentrated in sovereign and quasi-sovereign paper. Larcker and Watts (2020), in an early and widely-cited contribution on the US municipal market, argue that the greenium essentially disappears once green and conventional bonds of the same issuer are matched on issue date and maturity. The systematic review of MacAskill et al. (2021) likewise concludes that greenium estimates are highly sensitive to methodological choices. For sustainability-linked bonds the evidence is mixed: Koelbel and Lambillon (2023) identify a greenium-like effect of about 9 bps in the primary market, while Berrada et al. (2022) emphasise that the step-up structure of SL bonds generates option-like pricing features that are rarely internalised by the market, raising concerns about greenwashing.

Sector-specific work is still scarce. For the utility sector, Flammer (2020) and Tang and Zhang (2020) document positive stock-price reactions to green bond announcements. Fatica et al. (2021) refine these findings by showing that the greenium in the European corporate market is larger for issuers outside the financial sector, a pattern broadly consistent with utility bonds dominating the non-financial greenium. Earlier conceptual contributions such as Ehlers and Packer (2017) and the ongoing updates of the ICMA Green Bond Principles (ICMA, 2023) emphasise the role of external review and certification, while Agliardi and Agliardi (2021) provide a structural model showing that the greenium is theoretically consistent with a small but positive investor taste for green assets. Karpf and Mandel (2018) demonstrate that the sign of the greenium can switch over time even within a

single market segment, and Sangiorgi and Schopohl (2023) use survey evidence to argue that non-price motivations are often more important to issuers than any funding-cost advantage. At the sovereign level, Ando et al. (2023) estimate a sovereign greenium of only a few basis points. More recent analyses such as those by the ECB (2023) and Bloomberg NEF (2024) highlight the importance of controlling for rating, tenor and currency, which can create a composition bias in favour of labelled bonds. The present paper takes these concerns seriously and controls for such compositional features explicitly.

3. Hypotheses

Building on the literature and on the specific features of the utility sector, we test four hypotheses.

- *H1:* Utility-sector ESG-labelled bonds are associated with a statistically significant negative conditional spread differential, a greenium in the conventional sense, relative to otherwise comparable conventional utility bonds.
- *H2:* The unconditional greenium is substantially reduced or disappears once issuer composition (credit quality, tenor, issue size), benchmark, currency, region and time fixed effects are controlled for.
- *H3:* The greenium is larger for use-of-proceeds labels (green, social, sustainable) than for sustainability-linked bonds, because the latter carry additional step-up risk and are more exposed to greenwashing concerns.
- *H4:* The greenium is stronger in Western Europe than in North America and has compressed during the 2022 to 2026 tightening and partial-easing cycle relative to the 2020 to 2021 zero-rate period.

4. Data and methodology

We assemble a proprietary data set of all benchmark-sized utility-sector corporate bonds launched between 1 January 2020 and 15 April 2026. The raw data are sourced from a commercial new-issue database and cross-checked against the terms in the final offering circulars. The sector filter is Utilities and Power and covers electricity and gas, water and waste, renewable and alternative energy, nuclear, and related sub-segments. Private placements, retail tranches, and deals with a notional below USD 100 million equivalent are excluded. The final sample contains 2,725 tranches from 470 unique issuers across 47 countries, representing around USD 1,635 billion of aggregate notional. ESG-labelled issues account for 770 tranches (28.3% by count), of which 618 are green, 88 sustainability-linked, 63 sustainable, and 2 social. The dependent variable is the launch spread expressed in basis points over the relevant pricing benchmark. The benchmark type is included as a fixed effect, which preserves the information content of the spread while removing the mechanical differences across reference-rate conventions.

The variable of interest is an ESG indicator taking the value one whenever the deal is labelled green, social, sustainable or sustainability-linked, and zero otherwise. In an extended specification, we replace the composite indicator by three separate dummies for green, sustainable and sustainability-linked, in order to test Hypothesis 3. Control variables include: a numerical S&P rating (1 = AAA, 22 = D); tenor in years from the issue date to the scheduled maturity; the natural logarithm of the USD-equivalent issue size; and fixed effects for benchmark type, region of risk, currency of issue, and calendar year. We estimate the following cross-sectional model by ordinary least squares:

$$Spread_i = \alpha + \beta ESG_i + \gamma' X_i + \mu_t + \varepsilon_i$$

where: $Spread_i$ is the launch spread of tranche *i in bps*, ESG_i is the ESG dummy, X_i is the vector of deal-level numerical controls (rating, tenor, log size), μt captures year, region, currency and benchmark fixed effects, and ε_i is a residual.

Standard errors are heteroskedasticity-consistent (HC3). Identification rests on within-year, within-region, within-rating comparisons. The coefficient *b* is interpreted as the average spread differential paid by a utility ESG-labelled bond relative to an otherwise identical conventional tranche issued in the same year, region, currency and benchmark. A negative, statistically significant *b* provides evidence consistent with a greenium.

5. Results

Table 1 summarises the sample by ESG status. On a raw basis, ESG-labelled utility bonds price about 10 bps tighter than conventional peers (140.9 versus 151.2 bps). They are, however, also issued at shorter tenors (10.8 versus 13.0 years) and at slightly larger sizes, while carrying essentially identical average credit ratings. These univariate patterns already hint at a potential composition bias.

Table 1. Descriptive statistics by ESG status

Variable	ESG mean	Conv. mean	Difference	t-stat
Launch spread (bps)	140.9	151.2	−10.3	−2.29
Re-offer yield (%)	3.88	4.56	−0.68	−8.10
Issue size (USD m)	605.6	598.0	+7.6	0.60
Tenor (years)	10.8	13.0	−2.2	−6.50
S&P rating (num., 1=AAA)	8.22	8.25	−0.03	−0.29
Observations	770	1,955	–	–

Source: authors

Table 2 reports the OLS estimates of equation (1). Column (1) uses the composite ESG indicator. Column (2) disaggregates the indicator into green, sustainable and sustainability-linked dummies; the omitted category pools conventional and the two social bonds.

Table 2. OLS regression results. Dependent variable: launch spread (bps)

Variable	(1) Composite ESG	(2) Label-by-label
ESG (any label)	+12.44*** (3.48)	–
Green	–	+10.84*** (3.73)
Sustainable	–	+24.21*** (6.66)
Sustainability-linked	–	+9.19 (9.03)
S&P rating (num.)	+32.32*** (1.95)	+32.34*** (1.95)
Tenor (years)	+2.16*** (0.13)	+2.16*** (0.13)
Log USD issue size	−12.15** (5.31)	−11.99** (5.27)
Benchmark / Region / Currency / Year FE	Yes	Yes
Observations	2,020	2,020
Adjusted R^2	0.537	0.537

Source: authors

The regression sample falls from 2,725 tranches in the full universe to 2,020 observations because of essentially mechanical filters: a subset of tranches is missing a re-offer launch spread, a smaller subset is missing a usable senior rating, and a small number of perpetual structures are dropped. The excluded tranches do not differ systematically from the regression sample in their ESG share or regional composition. All specifications include benchmark-type, region, currency and year fixed effects.

Three results stand out. The composite ESG coefficient is positive and highly significant: all else equal, an ESG-labelled utility bond is associated with a launch spread about 12.4 bps wider than that of a comparable conventional bond. This is the opposite sign of the conventional greenium narrative. Second, disaggregating by label reveals that the conditional concession is driven by green (+10.8 bps) and especially by sustainable (+24.2 bps) bonds, while sustainability-linked bonds are not significantly different from conventional peers. Third, the controls behave as expected: each one-notch downgrade in the S&P rating is associated with about 32 bps more spread, each additional year of tenor with about 2.2 bps more, and larger issues pay modestly lower spreads. We use deliberately conditional language because the specification does not include issuer fixed effects or secondary-market-curve controls, and we therefore interpret the ESG coefficient as a conditional pricing association rather than as a causal pricing effect.

The sign reversal between the raw univariate comparison (ESG bonds 10 bps tighter) and the conditional estimate (ESG bonds 12 bps wider) is the most economically important finding. It demonstrates that the apparent unconditional greenium in utility-sector labelled bonds is almost entirely a composition artefact: ESG issues in this sector are concentrated in shorter tenors and in years with tighter secondary-market spreads, and once these effects are removed the pricing advantage not only vanishes but reverses. To test Hypothesis 4, we re-estimate column (1) separately on regional and temporal subsamples, reported in Table 3.

Table 3. ESG coefficient across subsamples

Subsample	N	ESG coeff. (bps)	p-value
Western Europe	553	+6.08	0.186
North America	1,329	+13.76**	0.013
Pre-2022 (low rates)	580	+8.88	0.187
2022 and later (tightening)	1,440	+8.93**	0.019
Investment grade only	1,903	+5.35*	0.068
High yield only	117	+23.59	0.464

Source: authors (Each row is a separate OLS regression with full controls and HC3 standard errors)

Three patterns emerge. First, the positive conditional concession is concentrated in North American issuance, where it reaches +13.8 bps and is significant at the 5% level, while in Western Europe the coefficient remains positive but smaller (+6.1 bps) and statistically indistinguishable from zero. This regional pattern is the mirror image of Hypothesis 4 and is consistent with a more mature and price-sensitive European sustainable-investor base. Second, the concession is stronger and statistically significant in the 2022-onwards period, the same window during which market participants reported a disappearing greenium. Third, the coefficient for the high-yield subsample is large but extremely imprecise, reflecting the thin tail of sub-investment-grade utility ESG issuance.

6. Discussion

The evidence does not support Hypothesis 1. There is no negative conditional pricing differential for utility-sector ESG-labelled bonds in the 2020 to 2026 primary market, and if anything the conditional sign is positive. Hypothesis 2 is supported: the small raw advantage is largely a composition artefact. Hypothesis 3 is partially supported, since sustainability-linked bonds do not outperform green bonds, but the largest concession is found on the sustainable label, which may reflect a thinner and more heterogeneous investor base for mixed-use-of-proceeds structures. Hypothesis 4 is rejected in its regional dimension but partially confirmed in its temporal dimension. Three non-exclusive mechanisms can rationalise these findings. First, supply dynamics: the utility sector has been one of the largest issuers of labelled debt, and the sheer volume of supply may have overwhelmed the price-insensitive dedicated-fund demand, especially in North America. Second, structural features: green bonds issued by utilities tend to fund projects, such as new renewable capacity or grid upgrades, that investors already price into the issuer conventional curve, reducing the marginal signalling value of the label. Third, an investor-base composition effect: the tightening cycle has re-opened the market for traditional real-money accounts for whom ESG certification is not a binding mandate, shifting marginal price formation away from dedicated green funds.

These three mechanisms reinforce rather than contradict one another. The supply channel and the investor-base channel are two sides of the same market-clearing condition: when a sector floods the primary market with labelled paper at the same time as the marginal buyer becomes a yield-driven real-money account rather than a mandate-constrained green fund, the relative scarcity value that underpins a conventional greenium is mechanically eroded. The utility sector is precisely where both forces are strongest, because it combines the largest labelled-issuance programmes in the corporate universe with capital-expenditure profiles already well understood by generalist credit investors. Seen in this light, a small positive conditional concession is the expected outcome of a maturing, well-supplied market segment in which the label no longer conveys scarce information. This reading is also consistent with the regional contrast: the European result sits close to zero, as expected where a deep dedicated-fund base still absorbs labelled supply, whereas the North American result is positive, as expected where that base is thinner.

For sustainable supply-chain finance the findings carry three practical implications. First, because the headline utility-sector greenium is largely a composition artefact, procurement decisions and long-dated supplier contracts that assume a structural funding-cost advantage from labelled debt may be over-optimistic. Second, the regional asymmetry suggests that energy supply chains anchored in North American utility issuers may face a higher marginal cost of labelled-debt financing for renewables, grid upgrades and storage. Third, the heterogeneity across sub-labels implies that the credibility of the sustainable and sustainability-linked categories deserves particular scrutiny from procurement managers and second-party reviewers.

Several caveats apply. First, we do not include issuer fixed effects; with 470 unique issuers, an issuer-FE specification would absorb a large share of the residual variation but would restrict identification to issuers that have brought both conventional and ESG-labelled deals to market, a non-trivially selected sub-population. Second, we do not control for the issuer outstanding secondary-market spread curve on the issue date. Third, we use a composite ESG indicator that pools four heterogeneous sub-labels. The term greenium is therefore used here as shorthand for the conditional pricing differential of the composite ESG indicator, not in the narrower sense of a use-of-proceeds green-bond premium. These extensions are left for future work.

7. Conclusion

Using a sample of 2,725 utility-sector bond tranches issued between January 2020 and April 2026, this paper tests whether ESG-labelled utility bonds, pooling green, social, sustainable and sustainability-linked structures, price at a differential relative to comparable conventional peers. Unconditionally, they appear to do so by around 10 bps; conditionally, once credit rating, tenor, issue size, benchmark, region, currency and year fixed effects are accounted for, the sign reverses and ESG-labelled bonds are associated with launch spreads approximately 12 bps wider than matched conventional peers. The positive conditional association is concentrated in North American issuance and in the post-2022 tightening cycle, while no statistically significant conditional differential is detected in Western Europe. The study therefore qualifies the widespread narrative of a universal utility-sector greenium. For treasurers, the results suggest that the funding-cost advantage of labelled issuance should not be taken for granted. For investors, the findings imply that utility ESG paper does not systematically under-compensate for credit risk in the current macro-financial regime. For regulators, standard-setters and procurement managers, the heterogeneous pricing across sub-labels underscores the importance of strengthening the credibility of the sustainable and sustainability-linked categories, where the pricing concession is largest.

References

Agliardi, E., & Agliardi, R. (2021). Corporate green bonds: Understanding the greenium in a two-factor structural model. Environmental and Resource Economics, 80(2), 257–278.

Ando, S., Fu, C., Roch, F., & Wiriadinata, U. (2023). How large is the sovereign greenium? IMF Working Paper 23/80, International Monetary Fund.

Baker, M., Bergstresser, D., Serafeim, G., & Wurgler, J. (2022). The pricing and ownership of US green bonds. Annual Review of Financial Economics, 14, 415–437.

Berrada, T., Engelhardt, L., Gibson, R., & Krueger, P. (2022). The economics of sustainability linked bonds. Swiss Finance Institute Research Paper 22-26.

Bloomberg NEF. (2024). Sustainable debt annual report 2024. Bloomberg New Energy Finance.

Caramichael, J., & Rapp, A. (2024). The green corporate bond issuance premium. Journal of Banking & Finance, 162, 107151.

Climate Bonds Initiative. (2024). Sustainable debt global state of the market 2023. London: CBI.

Ehlers, T., & Packer, F. (2017). Green bond finance and certification. BIS Quarterly Review, September 2017, 89–104.

European Central Bank. (2023). The pricing of green bonds in the euro area. Financial Stability Review, May 2023.

Fatica, S., Panzica, R., & Rancan, M. (2021). The pricing of green bonds: Are financial institutions special? Journal of Financial Stability, 54, 100873.

Flammer, C. (2020). Green bonds: Effectiveness and implications for public policy. Environmental and Energy Policy and the Economy, 1(1), 95–128.

Flammer, C. (2021). Corporate green bonds. Journal of Financial Economics, 142(2), 499–516.

International Capital Market Association. (2023). The green bond principles: Voluntary process guidelines for issuing green bonds. ICMA, Paris.

Karpf, A., & Mandel, A. (2018). The changing value of the green label on the US municipal bond market. Nature Climate Change, 8(2), 161–165.

Koelbel, J. F., & Lambillon, A.-P. (2023). Who pays for sustainability? An analysis of sustainability-linked bonds. Working paper, University of Zurich.

Larcker, D. F., & Watts, E. M. (2020). Where is the greenium? Journal of Accounting and Economics, 69(2–3), 101312.

MacAskill, S., Roca, E., Liu, B., Stewart, R. A., & Sahin, O. (2021). Is there a green premium in the green bond market? Systematic literature review revealing premium determinants. Journal of Cleaner Production, 280, 124491.

Pietsch, A., & Salakhova, D. (2022). Pricing of green bonds: Drivers and dynamics of the greenium. ECB Working Paper 2728, European Central Bank.

Sangiorgi, I., & Schopohl, L. (2023). Explaining green bond issuance using survey evidence: Beyond the greenium. The British Accounting Review, 55(1), 101071.

Tang, D. Y., & Zhang, Y. (2020). Do shareholders benefit from green bonds? Journal of Corporate Finance, 61, 101427.

Zerbib, O. D. (2019). The effect of pro-environmental preferences on bond prices: Evidence from green bonds. Journal of Banking & Finance, 98, 39–60.

LEAN AND RARE: ENVIRONMENTAL AWARD CRITERIA IN CENTRAL EUROPEAN TENDERING

Jakub Sieber, Erik Boltun

University of Economics in Bratislava, Slovakia
Faculty of Business Economics with seat in Kosice
jakub.sieber@euba.sk, erik.boltun@euba.sk

DOI: 10.35011/IDIMT-2026-323

Keywords

Public procurement; tendering; contract award; criteria; eForms; sustainability.

Abstract

Public procurement is increasingly expected to support public value, yet it remains unclear how far such ambitions are reflected in the design of tender evaluation. This paper examines how eForms public procurement notices structure award models in seven Central European countries: Austria, Czechia, Germany, Hungary, Poland, Slovakia and Slovenia. The study distinguishes lowest-price-only tenders from multi-criteria award models and examines the prevalence and weighting of environmental award criteria. Using 2024–2025 eForms data from Tenders Electronic Daily, the analysis covers 1,001,219 procurement lots and 1,639,200 criterion-level observations. Award-criterion types, names, descriptions and weights were extracted from XML notices and environmental criteria were classified using a conservative multilingual dictionary-based approach. The findings show that lowest-price-only evaluation remains dominant overall, country differences are substantial, and environmental award criteria are rare, appearing in only 1.5% of lots.

1. Introduction

Public procurement is a key mechanism through which public authorities translate budgets into market demand. It shapes what governments buy, how suppliers compete, which offer attributes are rewarded and how public value is operationalised in purchasing decisions. Procurement therefore acts as a strategic interface between public authorities and supplier markets (Grandia & Meehan, 2017). In the European Union, this role is reinforced by a legal framework that allows contracts to be awarded not only on price, but also on the most economically advantageous tender, including qualitative, environmental and other criteria linked to the contract subject matter (European Parliament and Council of the European Union, 2014). A persistent tension concerns the role of price. Price-only award models are simple, transparent and easy to administer. They also limit discretion, which can be attractive in regulated settings where accountability and comparability are important (Bosio et al., 2022). However, price-only evaluation captures a narrow view of value because it focuses on direct purchase price and leaves little explicit space for life-cycle considerations or environmental performance (Dotoli et al., 2020). Multi-criteria award models offer an alternative by combining price with one or more non-price criteria. Yet their practical meaning depends on the content and weighting of the criteria used. A tender with a marginally weighted non-price criterion

may remain price-dominant in practice, whereas substantial weights assigned to technical, quality or environmental criteria can alter the competitive logic of the award decision (Dotoli et al., 2020). This paper examines environmental award criteria (EAC) within procurement award models. Sustainable public procurement is often discussed through overlapping labels, including green, sustainable, circular, socially responsible and strategic procurement (Brammer & Walker, 2011; Cheng et al., 2018). Although these concepts are important, treating sustainability as a broad umbrella can introduce conceptual ambiguity in large-scale procurement notice data. This study therefore focuses on EAC as an observable text-based category in eForms tender notices. This focus is consistent with public procurement research that treats environmental criteria as one mechanism through which public buyers integrate environmental performance into tender evaluation (Testa et al., 2016). The distinction between award-model structure and environmental criterion use is central. A buyer may formally move beyond price by adopting a multi-criteria model while still assigning most evaluative weight to price. The empirical question is therefore whether environmental considerations are visible in tender evaluation structures, how often they appear and what weight they receive. This distinction between legal possibility and observed award-model practice is essential for assessing whether environmental objectives are embedded in competitive evaluation or remain peripheral to award decisions (Behravesh et al., 2022). Since October 2023, eForms have been mandatory for publishing EU public procurement notices, replacing older forms with a more structured digital reporting regime.

Accordingly, this paper asks: *To what extent do eForms public procurement notices rely on price-only versus multi-criteria award models, and how are evaluation weights distributed across award-criterion categories?*

2. Theoretical background

Public procurement award decisions reflect how contracting authorities define value in a specific tender. Before a contract is awarded, the buyer must decide which offer attributes will matter, how these attributes will be assessed, and how price will be balanced against other dimensions of performance. The award model therefore operates as the evaluative architecture of the tender: it translates procurement objectives into ranking rules and determines the basis on which suppliers compete. This distinction matters because public procurement is increasingly understood as a policy instrument. Through procurement, public authorities can shape markets, stimulate innovation and pursue broader public objectives, including sustainability and environmental protection (Grandia & Meehan, 2017). The EU procurement framework allows contracting authorities to award contracts based on the most economically advantageous tender, this may be identified through price, cost or the best price-quality ratio, including environmental and other aspects linked to the subject of the contract (European Parliament, 2014). A multi-criteria model is one in which price is combined with at least one additional criterion, such as quality, technical merit, or environmental performance. Supplier selection depends on the scoring and weighting rules through which price and non-price attributes are converted into a final ranking (Watt et al., 2009). These rules shape the competitive logic of the tender. If price receives all evaluative weight, suppliers compete primarily on cost. If non-price criteria receive meaningful weight, suppliers can compete on additional dimensions.

Environmental criteria are a specific form of non-price evaluation. Public procurement has long been viewed as a tool for advancing environmental objectives, particularly through green public procurement, understood as purchasing goods, services and works with reduced environmental impacts across their life cycle (European Commission, 2016). Prior research shows that green public procurement can support environmental policy implementation, stimulate greener markets and encourage suppliers to develop more sustainable products and processes (Krieger & Zipperer, 2022). However, sustainable, green, circular, responsible and strategic procurement are sometimes used

interchangeably, although they capture different empirical content. Treating sustainability as a single category may therefore conflate criteria as diverse as emissions, recyclability, employment of disadvantaged workers, working conditions, community benefits and ethical sourcing. In this paper, EAC are understood as award criteria referring to the environmental performance or life-cycle impact of the procured good, service or work. This includes energy efficiency, emissions, waste reduction, recyclability, circularity, resource use, environmental management and related green-performance attributes (Testa et al., 2016). Because weighted EAC can advantage suppliers offering better environmental performance, the award stage is a relevant place to observe whether environmental value enters supplier competition. Their assigned weight is as important as their presence: a low-weight EAC may signal environmental awareness while having limited influence on supplier ranking.

3. Methodology

3.1. Research design, data and empirical scope

This study uses a document-based research design to examine how contracting authorities specify award criteria in public procurement notices. Such notices provide a comparable administrative record of award-model design before bid evaluation. The analysis distinguishes lowest-price-only tenders from multi-criteria tenders that combine price with at least one additional award criterion. This distinction reflects the institutional logic of EU public procurement, where contracts may be awarded on price alone or on other criteria linked to the subject matter of the contract (European Parliament and Council of the European Union, 2014). The data are drawn from eForms notices published in Tenders Electronic Daily (TED), the official EU platform for public procurement notices. eForms constitute the current structured notice standard established under Regulation (EU) 2019/1780. The study covers notices published in 2024 and 2025 in seven Central European countries: Austria (AT), Czechia (CZ), Germany (DE), Hungary (HU), Poland (PL), Slovenia (SI) and Slovakia (SK). These countries share the EU procurement framework but differ in administrative practice and in their use of price-only and multi-criteria award models. The unit of analysis is the procurement lot. Lot-level analysis is appropriate because award criteria are typically specified for individual lots rather than only for the notice. Where a notice contained multiple lots, each lot was treated as a separate observation. The final sample comprises 1,001,219 lot-level observations and 1,639,200 criterion-level observations.

Table 1. Sample distribution by country

Country	Lots 2024	Lots 2025	Total lots	Share of sample (%)
Poland	292,278	327,123	619,401	61.86
Germany	116,359	112,597	228,956	22.87
Slovenia	24,251	27,496	51,747	5.17
Czechia	22,328	26,780	49,108	4.90
Hungary	12,233	12,872	25,105	2.51
Slovakia	7,361	8,889	16,250	1.62
Austria	5,393	5,259	10,652	1.06
Total	480,203	521,016	1,001,219	100.00

Source: authors

3.2. Data extraction, environmental award-criterion classification and statistics

The extraction procedure parsed eForms XML files and retained notices where the buyer country could be identified from contracting-party organisational address information. At criterion level, the dataset retained award-criterion type, name, description, weight parameter code, numeric weight value, CPV8 code, source file, lot identifier, notice identifier, buyer country and year. Percentage weights were used directly when explicitly coded and when lot-level weights summed to approximately 100. For single price-only lots without numeric weight, the price weight was imputed as 1.00. Numeric values not explicitly coded as percentages were recovered only when all criterion values within a lot summed to 100 and were not fixed amounts, thresholds or order-of-importance values. Ambiguous or incomplete values were excluded from the analysis. EAC were identified at criterion level using a conservative multilingual dictionary-based classification implemented in Python. The dictionary covered environmental, ecological, emissions, energy efficiency, recycling, reuse, circularity, environmental management and equivalent terms in the sample languages. Dictionary rules are appropriate when the construct is theoretically specified, the coding logic must remain transparent, and false positives and false negatives can be audited (Krippendorff, 2018). A criterion was coded as environmental only when it directly evaluated environmental performance, impact or comparable performance attributes. Criteria referring only to personnel, logistics, response time, or generic service quality were excluded unless they contained clear environmental evidence.

Environmental award criteria were identified using a conservative multilingual dictionary-based classification implemented at criterion level. The dictionary covered the languages represented in the country sample, namely German, Czech, Hungarian, Polish, Slovak and Slovene, with English variants retained where they appeared in criterion text. The classification was applied to the criterion name and description, together with the reported award-criterion type. The dictionary included terms referring to environmental performance, ecological performance, emissions, CO_2, greenhouse gases, energy efficiency, energy consumption, recycling, recycled content, reuse, waste reduction, circularity, resource efficiency, life-cycle performance, environmental management systems, ISO 14001, EMAS and comparable green-performance attributes. A criterion was coded as environmental only where the wording indicated that environmental performance, environmental impact or life-cycle-related performance formed part of the tender evaluation.

Table 2. Examples of included and excluded dictionary terms

Classification decision	Examples	Coding logic
Included	energy efficiency; energy consumption; CO_2 emissions; greenhouse gas emissions; low-emission vehicles; recycling; reuse; circularity; waste reduction; ISO 14001; EMAS; life-cycle impact	Included when the criterion evaluates environmental performance, environmental impact or life-cycle-related attributes
Excluded	delivery time; response time; staff experience; warranty; technical quality; project organisation; logistics plan; general service quality	Excluded because these criteria do not directly evaluate environmental performance
Excluded unless environmental content is explicit	green area maintenance; landscaping; cleaning quality; transport organisation; sustainability strategy	Included only when the criterion explicitly evaluates environmental outcomes, such as emissions, waste, resource use, biodiversity, etc.

Source: authors

Classification reliability was assessed through a stratified manual validation of 300 criterion-level records. The validation sample included 100 records classified by the dictionary as environmental, 100 difficult non-environmental records containing environmental-looking vocabulary, and 100 general non-environmental non-price records. Each record was manually checked against the criterion name, description and award-criterion type, following the same definition of environmental performance used in the automated classification. The validation produced 89 true positives, 11 false positives, 163 true negatives and 37 false negatives. This corresponds to precision of 0.890, recall of 0.706, specificity of 0.937 and an unweighted validation-sample accuracy of 0.840.

The descriptive analysis reports counts and percentages of award models and EAC by country and year. The main inferential comparison tests whether award-model composition differs by country within comparable CPV8 categories. For each eligible CPV8 category, a buyer-country-by-award-model contingency table was constructed and tested using a chi-square test of independence, which is appropriate for categorical variables. Given the large sample size, statistical significance alone is not treated as substantively informative. Therefore, p-values are interpreted together with Cramér's V and the cross-country range in multi-criteria shares, which indicate the size and practical relevance of within-CPV8 differences. As an exploratory extension, non-price weight shares are compared across countries within eligible CPV8 categories among multi-criteria lots with usable weights. Because these shares are bounded between 0 and 1, the Kruskal-Wallis rank-sum test is used for distributional comparisons (Kruskal and Wallis, 1952).

4. Findings

Across the sample, 56.71% of lots rely on lowest-price-only evaluation, 39.79% use multi-criteria evaluation with price, and 3.50% fall into the other or unclear category. However, this aggregate pattern hides substantial cross-country variation. SK is the most price-oriented case, while AT is the clearest contrasting case. Only 19.08% of AT lots are classified as lowest-price-only, while 72.67% use multi-criteria evaluation. Table 3 reports the descriptive comparison between lowest-price-only and multi-criteria award models.

Table 3. Award criteria composition by country

Country	Lowest-price-only lots	Multi-criteria lots	Other/ unclear lots	Lowest-price-only (%)	Multi-criteria (%)	Other/ unclear (%)
Austria	2,032	7,741	879	19.08	72.67	8.25
Czechia	36,326	10,334	2,448	73.97	21.04	4.98
Germany	134,847	71,053	23,056	58.90	31.03	10.07
Hungary	13,224	11,544	337	52.67	45.98	1.34
Poland	330,881	281,614	6,906	53.42	45.47	1.11
Slovenia	35,565	15,094	1,088	68.73	29.17	2.10
Slovakia	14,890	1,038	322	91.63	6.39	1.98
Total	567,765	398,418	35,036	56.71	39.79	3.5

Source: authors

The final audited dataset identifies 15,033 environmental lots out of 1,001,219 lots. This corresponds to an overall environmental prevalence of approximately 1.50%. EAC are therefore not a routine feature of eForms award evaluation in the sample. Table 4 reports price and non-price weights for

lots where weight information can be meaningfully summarised. The results show that multi-criteria evaluation does not necessarily imply low price dominance. In AT, SI, and SK, the median price weight is 0.80. This means that even where price is combined with other criteria, the median tender still gives 80% of the evaluative weight to price and only 20% to non-price criteria.

Table 4. Environmental award criteria by country

Country	Lots *n*	Median price weight	Mean price weight	Median non-price weight	Mean non-price weight	Median quality weight	Mean quality weight
Austria	4,453	0.80	0.711	0.20	0.289	0.20	0.289
Czechia	9,750	0.75	0.728	0.25	0.273	0.25	0.272
Germany	48,468	0.50	0.496	0.50	0.506	0.50	0.505
Hungary	101	0.40	0.482	0.60	0.518	0.60	0.518
Poland	150,962	0.60	0.662	0.40	0.338	0.40	0.337
Slovenia	1,168	0.80	0.783	0.20	0.217	0.20	0.214
Slovakia	257	0.80	0.735	0.20	0.265	0.20	0.267

Source: authors

The CPV8 descriptive results show that the mix of award models differs substantially even within common procurement categories. Table 5 reports selected high-volume CPV8 categories. The variation is large. Pharmaceutical products are strongly price-oriented, with 83.75% lowest-price-only and only 15.57% multi-criteria lots. By contrast, architectural, construction, engineering and inspection services are predominantly multi-criteria, with 68.85% multi-criteria lots and only 16.92% lowest-price-only. IT services also lean strongly towards multi-criteria evaluation, with 66.07% multi-criteria lots. Engineering services show a similar pattern, with 65.44% multi-criteria lots.

Table 5. Selected high-volume CPV8 categories by award-model composition

CPV8	Lots	Lowest-price-only (%)	Multi-criteria (%)	Environmental (%)	Median number of criteria
Pharmaceutical products	116,883	83.75	15.57	0.06	1
Medical consumables	68,093	44.84	53.26	0.04	2
Medical equipment	35,151	33.22	64.70	1.91	2
Construction work	17,073	67.16	31.10	1.38	1

Source: authors

Standardised goods, such as pharmaceuticals, show high use of lowest-price-only evaluation. More complex services, such as engineering, architectural, and IT services, show higher use of multi-criteria evaluation. Table 6 reports the CPV8 fixed-effects robustness model for multi-criteria use, with DE as the reference country. The model estimates country differences in the probability of using multi-criteria evaluation while comparing lots within CPV8 categories. Relative to DE, AT is 37.53 percentage points more likely to use multi-criteria evaluation within comparable CPV8 categories. By contrast, SK is 26.41 percentage points lower than DE.

Table 6. Country differences in multi-criteria use relative to Germany, with CPV8 fixed effects

Country	Coefficient, pp	SE, pp	95% CI low	95% CI high	p-value
Austria	37.53	0.66	36.25	38.82	<0.001
Czechia	-10.78	0.30	-11.36	-10.20	<0.001
Hungary	11.69	0.44	10.82	12.56	<0.001
Poland	25.25	0.20	24.86	25.65	<0.001
Slovenia	-8.46	0.29	-9.02	-7.89	<0.001
Slovakia	-26.41	0.39	-27.17	-25.64	<0.001

Source: authors

In addition, Table 7 reports country differences in EAC relative to Germany. Austria is 1.36 percentage points higher, while CZ, PL, and SI are significantly lower. HU and SK do not differ significantly from DE in this model.

Table 7. Country differences in environmental criteria relative to Germany

Country	Coefficient, pp	SE, pp	95% CI low	95% CI high	p-value
Austria	1.36	0.27	0.83	1.88	<0.001
Czechia	-0.67	0.09	-0.84	-0.50	<0.001
Hungary	0.13	0.13	-0.13	0.39	0.3299
Poland	-0.62	0.07	-0.76	-0.47	<0.001
Slovenia	-0.66	0.09	-0.84	-0.49	<0.001
Slovakia	0.08	0.15	-0.21	0.37	0.5951

Source: authors

5. Conclusion

To conclude, the findings provide a coherent picture of award-model design in the audited eForms data. *First*, lowest-price-only evaluation remains the most common award model in the full sample, but its use differs sharply across countries. Slovakia, Czechia, Slovenia, and Germany are more price-oriented, while Austria is strongly multi-criteria and Poland and Hungary occupy intermediate positions. *Second*, EAC are rare. They appear in only 1.5% of all lots. The findings therefore do not support a broad claim that EAC have become a routine feature of public procurement evaluation under eForms. *Third*, the weight analysis shows that multi-criteria incidence and substantive non-price importance are not the same. Some countries use multi-criteria models frequently but keep price highly dominant. *Fourth*, CPV8 robustness checks show that the observed country differences are not merely a result of different countries buying different objects. Country effects remain visible within common CPV8 categories, and the fixed-effects results show large differences. The same conclusion applies to non-price weights: within comparable procurement categories, countries differ substantially in how much evaluative weight they assign to non-price criteria.

Overall, the findings support the central argument of the paper. The eForms data do not simply show whether public buyers use “MEAT” or “lowest price”. They reveal the evaluative architecture of procurement: whether tenders are price-only or multi-criteria, which non-price criteria appear, and

how much weight those criteria receive. This architecture differs markedly across countries and procurement categories. EAC remain rare and lean.

References

Behravesh, S.-A., Darnall, N., & Bretschneider, S. (2022). A framework for understanding sustainable public purchasing. Journal of Cleaner Production, 376, 134122. https://doi.org/10.1016/j.jclepro.2022.134122

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. Journal of the Royal Statistical Society: Series B, 57(1), 289-300.

Bosio, E., Djankov, S., Glaeser, E. L., & Shleifer, A. (2022). Public procurement in law and practice. American Economic Review, 112(4), 1091–1117. https://doi.org/10.1257/aer.20200738

Brammer, S., & Walker, H. (2011). Sustainable procurement in the public sector: An international comparative study. International Journal of Operations & Production Management, 31(4), 452–476. https://doi.org/10.1108/01443571111119551

Cheng, W., Appolloni, A., D'Amato, A., & Zhu, Q. (2018). Green public procurement, missing concepts and future trends: A critical review. Journal of Cleaner Production, 176, 770–784. https://doi.org/10.1016/j.jclepro.2017.12.027

Dotoli, M., Epicoco, N., & Falagario, M. (2020). Multi-Criteria Decision Making techniques for the management of public procurement tenders: A case study. Applied Soft Computing, 88, Article 106064. https://doi.org/10.1016/j.asoc.2020.106064

European Commission. (2016). Buying green! A handbook on green public procurement (3rd ed.). Publications Office of the European Union. https://doi.org/10.2779/837689

European Parliament (2014). Directive 2014/24/EU of the European Parliament and of the Council of 26 February 2014 on public procurement. Official Journal of the European Union, L 94, 65–242.

Fazekas, M., & Tóth, I. J. (2016). From corruption to state capture: A new analytical framework with empirical applications from Hungary. Political Research Quarterly, 69(2), 320-334.

Grandia, J., & Meehan, J. (2017). Public procurement as a policy tool: Using procurement to reach desired outcomes in society. International Journal of Public Sector Management, *30*(4), 302–309. https://doi.org/10.1108/IJPSM-03-2017-0066

Krieger, B., & Zipperer, V. (2022). Does green public procurement trigger environmental innovations? Research Policy, 51(6), 104516. https://doi.org/10.1016/j.respol.2022.104516

Krippendorff, K. (2018). Content Analysis: An Introduction to Its Methodology. 4th ed. Thousand Oaks, CA: SAGE.

Kruskal, W. H., & Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis. Journal of the American Statistical Association, 47(260), 583-621.

Testa, F., Annunziata, E., Iraldo, F., & Frey, M. (2016). Drawbacks and opportunities of green public procurement: An effective tool for sustainable production. Journal of Cleaner Production, 112, 1893–1900. https://doi.org/10.1016/j.jclepro.2014.09.092

Watt, D. J., Kayis, B., & Willey, K. (2009). Identifying key factors in the evaluation of tenders for projects and services. International Journal of Project Management, 27(3), 250–260. https://doi.org/10.1016/j.ijproman.2008.03.002

WHEN ONE OFFER IS ALL THERE IS: PROCUREMENT SAVINGS IN OBSERVED SINGLE-OFFER TENDERS

Michal Tkáč, Michal Tkáč, Alžbeta Džujková

Bratislava University of Economics and Business, Slovakia
Faculty of Business Economics with seat in Kosice
michal.tkac1@euba.sk, michal.tkac@euba.sk, alzbeta.dzujkova@euba.sk

DOI: 10.35011/IDIMT-2026-331

Keywords

Public procurement; single-bid procurement; procurement savings; tender competition; contract award notices; TED.

Abstract

Competition is widely regarded as a key determinant of procurement efficiency, yet research on single-bid procurement has focused primarily on its incidence rather than its consequences. This study examines how procurement savings differ between observed single-offer and observed multiple-offer tenders using tender-level data derived from Tenders Electronic Daily (TED) Contract Award Notices covering procurement years 2018–2023. The analysis focuses on tenders with observed offer-count information and measures procurement savings as the percentage difference between estimated and final award values. The results show that observed multiple-offer tenders achieve substantially higher procurement savings than observed single-offer tenders. The observed difference remained statistically significant across all reported analyses and persists within electronic auctions, framework agreements, and all 45 examined CPV divisions. Electronic auctions are associated with moderately higher savings, whereas framework agreements are associated with lower savings, particularly among observed multiple-offer tenders. The magnitude of the difference varies considerably across procurement markets: five CPV divisions exhibit median savings differences exceeding five percentage points, whereas sixteen divisions exhibit differences of one percentage point or less. The findings suggest that limited observed competition is consistently associated with weaker price movement and lower procurement savings, highlighting the economic consequences of observed competition in public procurement.

1. Introduction

Public procurement is one of the most important mechanisms for allocating public resources. In OECD countries, it accounts for 10–15% of GDP annually (Bosio et al., 2022). Effective allocation requires competition because competition encourages suppliers to submit bids closer to their actual costs and reduces unnecessary public expenditure. Vickrey (1961) showed that as the number of bidders increases, bids move systematically toward true costs, while Bulow and Klemperer (1996) demonstrated that attracting an additional bidder often generates greater benefits than improvements in negotiation. Consequently, limited competition may reduce procurement efficiency and weaken price discipline.

2. Literature review

Economic theory and empirical evidence suggest that competition improves procurement outcomes. Matilla-García and Vega (2024) estimated that each additional bidder reduces contract prices by approximately 2.1%, while Carril, Gonzalez-Lira, and Walker (2026) found that expanding the pool of informed bidders lowers procurement costs, particularly in markets with previously limited competition. Similar evidence has been reported for electronic reverse auctions, where competitive conditions were associated with higher procurement savings (Dráb et al., 2022). Competitive environments also influence bidding behaviour through mechanisms such as the winner's curse (Thaler, 1988). Together, these studies suggest that reduced competition should be associated with weaker price movement and lower procurement savings.

Observed single-offer outcomes do not necessarily imply the existence of only one capable supplier. Suppliers participate only when expected benefits exceed participation costs (McAfee & McMillan, 1987), while contractual requirements and procurement complexity may discourage participation (Drake et al., 2025). Greater transparency has been shown to increase bidder participation and bidding aggressiveness (Coviello & Mariniello, 2014; De Silva et al., 2008). Recent studies have examined competition through bidder participation, contract prices, procurement savings, and the incidence of single-bid procurement (Fazekas & Blum, 2021; Tátrai et al., 2024; Titl, 2023). However, less attention has been paid to procurement savings among tenders with observed offer-count information.

This study addresses that gap by comparing procurement savings between observed single-offer and observed multiple-offer tenders using a transparent TED-based tender-level dataset. The study further examines whether the relationship between observed competition and procurement savings varies across procurement settings and procurement categories.

3. Methodology

The study uses procurement data from the sample derived from Tenders Electronic Daily (TED), the official publication platform for European public procurement notices. The sample covers procurement years 2018–2023 and uses Contract Award Notices (ID_NOTICE_CAN) as the unit of analysis. Each observation represents a tender-level notice. Delina and Macik (2023) emphasize the importance of data quality and variable construction in procurement analytics, which is particularly relevant when transforming administrative procurement records into analytical datasets. The key outcome variable is procurement savings, calculated as the percentage difference between estimated and final award values: tender_saving_pct = ((estimated value − final value) / estimated value) × 100. Award values were first deduplicated using the combination of ID_NOTICE_CAN and ID_AWARD and subsequently aggregated to the tender level. Only tenders with positive and complete estimated and final values were retained. Tenders were classified as observed single-offer when at least one record reported NUMBER_OFFERS = 1 and no record reported more than one offer. Tenders were classified as observed multiple-offer when at least one record reported NUMBER_OFFERS > 1. Tenders with missing or unusable offer-count information were excluded because they could not be reliably assigned to either group. Finally, observations with procurement savings outside the interval [-50%, 50%] were removed to reduce the influence of reporting errors and extreme observations. The final analytical sample consists of 194,571 observed single-offer tenders and 346,032 observed multiple-offer tenders, resulting in a total sample of 540,603 tenders.

4. Research results

4.1. Procurement savings and number of offers

The primary objective of this study is to compare procurement savings between observed single-offer and observed multiple-offer tenders. Because procurement savings were not normally distributed in either group (D'Agostino–Pearson K^2, $p < 0.001$), nonparametric methods were used throughout the analysis. Figure 1 summarizes the differences in procurement savings between the two offer groups.

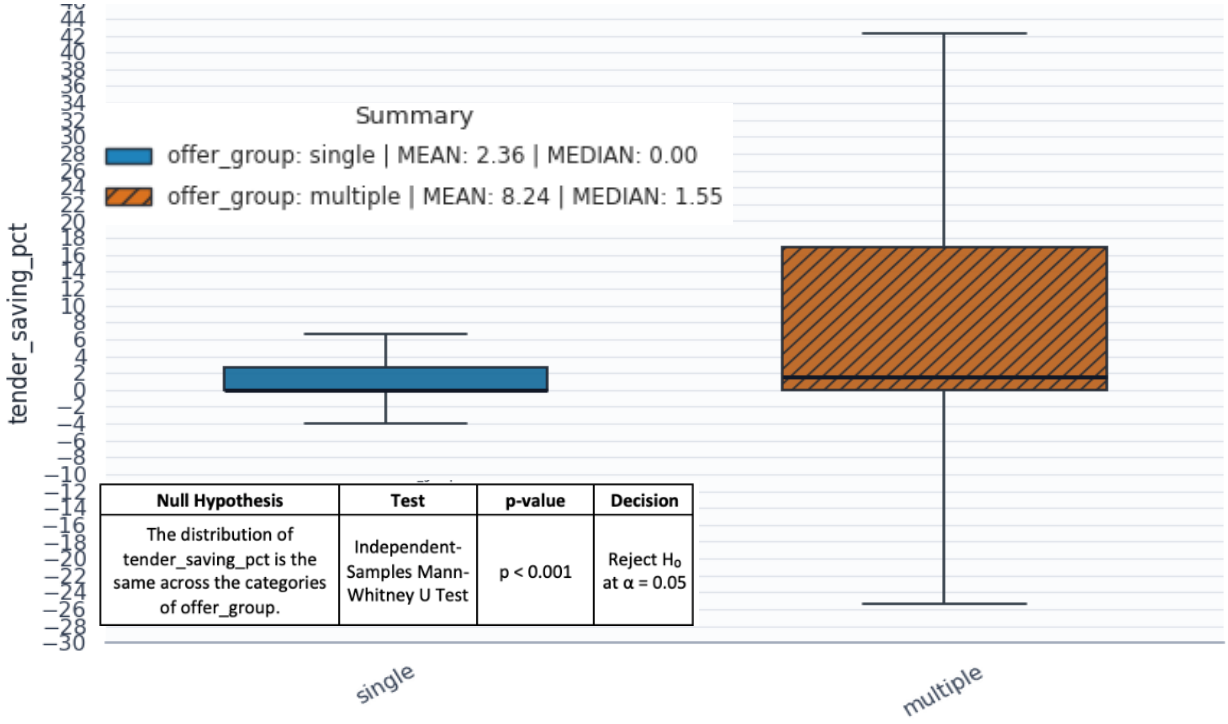


Figure 1. Procurement savings by offer group: Boxplot comparison and Mann–Whitney U test results

Source: authors

Observed multiple-offer tenders achieved a mean saving of 8.24%, compared with 2.36% for observed single-offer tenders. The median saving was 1.55% for observed multiple-offer tenders but 0.00% for observed single-offer tenders, indicating that at least half of the observed single-offer tenders generated no measurable saving relative to the estimated contract value. The interquartile range for observed multiple-offer tenders extended from 0.00% to 16.89%, whereas the corresponding range for observed single-offer tenders was substantially narrower, from 0.00% to 2.67%. Because the normality assumption was rejected for both groups in the previous section and variance homogeneity was also rejected (Levene statistic = 25,091.55, $p < 0.001$), group differences were evaluated primarily using the Mann–Whitney U test. The test indicated a statistically significant difference between the distributions of procurement savings for observed single-offer and observed multiple-offer tenders ($U = 26{,}437{,}973{,}278.5$, $p < 0.001$). The associated rank-biserial effect size was -0.21, suggesting a meaningful difference in procurement outcomes between the two offer groups. A supplementary Mood median test reached the same conclusion ($\chi^2 = 10{,}859.37$, $p < 0.001$), confirming that the difference is not limited to mean values but is also reflected in the central tendency of the distributions. The results indicate that observed multiple-offer tenders are associated with substantially higher procurement savings than observed single-offer tenders. This pattern is evident in both the descriptive statistics and the nonparametric comparison of the two distributions.

4.2. Procurement savings and electronic auctions

To further examine the relationship between competition and procurement outcomes, procurement savings were analysed jointly by offer group and the use of electronic auctions. This comparison allows assessment of whether electronic auctions are associated with higher procurement savings within both observed single-offer and observed multiple-offer tenders. Figure 2 presents boxplots of procurement savings for the four resulting categories together with nonparametric tests of distributional and median differences.

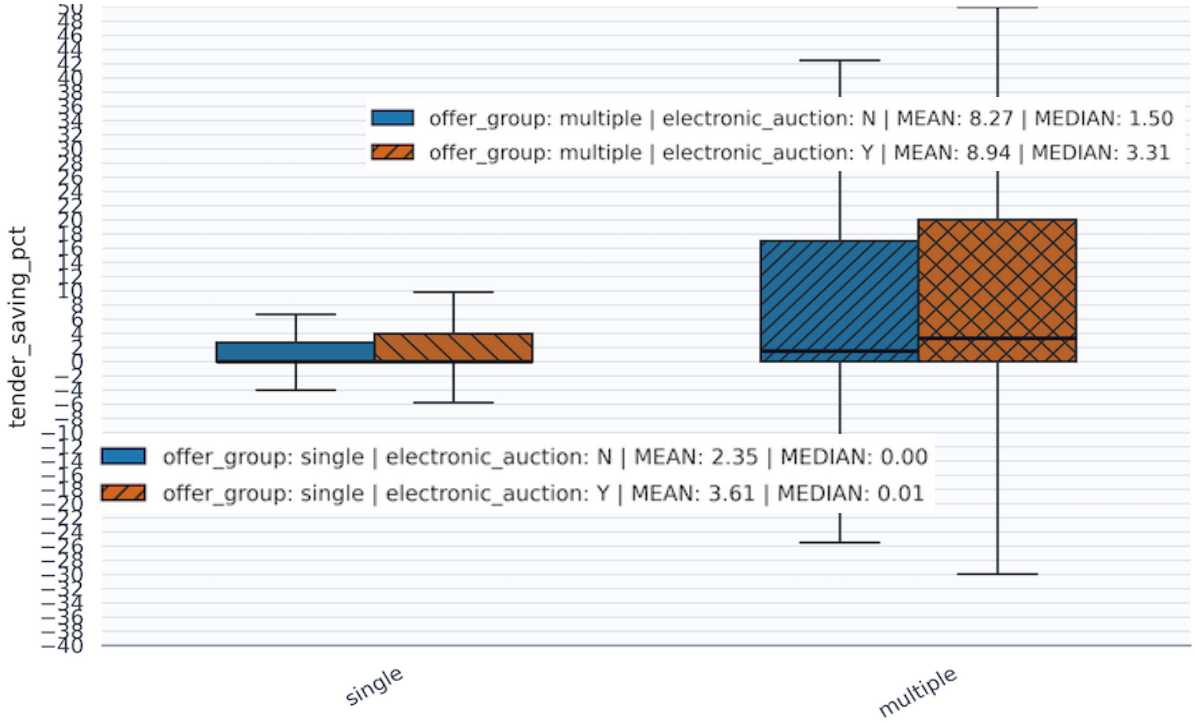


Figure 2. Boxplot comparison of procurement savings and electronic auctions

Source: authors

Among observed single-offer tenders, procurement savings remained relatively low regardless of auction usage. Single-offer tenders without electronic auctions (N = 187,471) achieved a mean saving of 2.35% and a median saving of 0.00%. The interquartile range extended from 0.00% to 2.67%. In comparison, single-offer tenders using electronic auctions (N = 2,224) achieved a mean saving of 3.61% and a median saving of 0.01%, with an interquartile range from 0.00% to 3.93%. The Mann–Whitney U test indicated a statistically significant difference between the distributions (U = 190,373,062.5, $p < 0.001$), although the associated rank-biserial effect size was relatively small (-0.09). The Mood median test likewise rejected the null hypothesis of equal medians ($\chi^2 = 64.31$, $p < 0.001$). A similar pattern was observed among observed multiple-offer tenders. Multiple-offer tenders without electronic auctions (N = 327,766) achieved a mean saving of 8.27% and a median saving of 1.50%, with an interquartile range from 0.00% to 17.00%. Multiple-offer tenders using electronic auctions (N = 8,585) achieved a mean saving of 8.94% and a median saving of 3.31%, while the interquartile range extended from 0.00% to 20.00%. The Mann–Whitney U test again identified a statistically significant difference between the distributions (U = 1,340,770,889, $p < 0.001$), although the corresponding rank-biserial effect size was modest (-0.05). The Mood median test likewise rejected the null hypothesis of equal medians ($\chi^2 = 81.17$, $p < 0.001$). Comparing all four categories reveals that the distinction between observed single-offer and observed multiple-offer tenders is substantially larger than the difference associated with auction usage alone. Electronic auctions were associated with higher procurement savings in both offer groups ($p < 0.001$). Among observed

multiple-offer tenders, median savings increased from 1.50% to 3.31% when electronic auctions were used, whereas savings among observed single-offer tenders remained close to zero regardless of auction usage. The results indicate that procurement savings are more strongly associated with the presence of multiple offers than with auction usage itself.

4.3. Procurement savings and framework agreements

Framework agreements represent a commonly used procurement mechanism intended to simplify repeated purchasing and reduce transaction costs. To examine whether procurement savings differ between framework and non-framework procurement, savings were analyzed separately for observed single-offer and observed multiple-offer tenders. Figure 3 presents the distributions of procurement savings across these categories together with nonparametric tests of distributional and median differences.

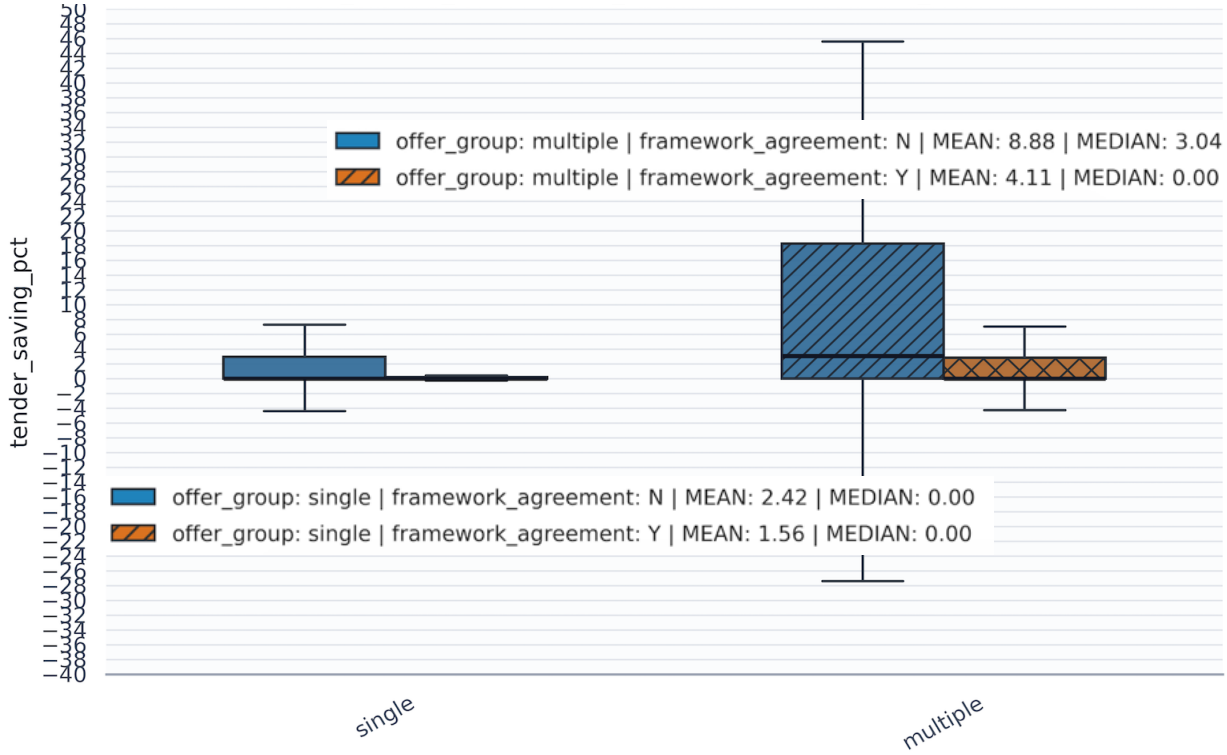


Figure 3. Boxplot comparison of procurement savings and framework agreements

Source: authors

Among observed multiple-offer tenders, substantial differences were observed between framework and non-framework procurement. Multiple-offer tenders outside framework agreements (N = 299,793) achieved a mean saving of 8.88% and a median saving of 3.04%, with an interquartile range extending from 0.00% to 18.26%. In contrast, multiple-offer tenders conducted through framework agreements (N = 46,239) achieved a mean saving of 4.11% and a median saving of 0.00%, with an interquartile range from 0.00% to 2.83%. The Mann–Whitney U test indicated a statistically significant difference between the distributions ($U = 8{,}741{,}462{,}006$, $p < 0.001$), with a rank-biserial effect size of 0.26. The Mood median test likewise rejected the null hypothesis of equal medians ($\chi^2 = 11{,}880.55$, $p < 0.001$). A similar pattern was observed among observed single-offer tenders, although the differences were smaller. Single-offer tenders outside framework agreements (N = 179,662) achieved a mean saving of 2.42%, compared with 1.56% for framework agreements (N = 14,909). Median savings were 0.00% in both categories. However, the interquartile range was notably narrower for framework agreements (0.00%–0.17%) than for non-framework procurement (0.00%–

2.91%). The Mann–Whitney U test again identified a statistically significant difference between the distributions (U = 1,502,158,742, $p < 0.001$), although the corresponding effect size was smaller (0.12). The Mood median test also rejected the null hypothesis of equal medians ($\chi^2 = 1{,}412.79$, $p < 0.001$). The comparison indicates that framework agreements are associated with lower procurement savings in both offer groups. The difference is particularly pronounced among observed multiple-offer tenders, where median savings decline from 3.04% outside framework agreements to 0.00% within framework agreements. The stronger effect size observed in the multiple-offer group further suggests that the relationship between framework agreements and procurement savings is more substantial when multiple offers are observed.

4.4. Heterogeneity of procurement savings across procurement categories

To examine whether the relationship between observed competition and procurement savings varies across procurement markets, separate analyses were conducted for 45 CPV divisions. Table 1 summarizes the magnitude of the median savings difference between observed multiple-offer and observed single-offer tenders across procurement categories.

Table 1. CPV Divisions grouped by the magnitude of median savings differences between observed Multiple-Offer and observed Single-Offer tenders

Magn.	CPV divisions (median difference; one-bid share)
Large (>5 pp)	Architectural, construction, engineering and inspection services (8.80; 20.4%), Collected and purified water (7.84; 86.8%), Agricultural, forestry, horticultural, aquacultural and apicultural services (6.07; 45.2%), Sewage, refuse, cleaning and environmental services (5.79; 32.0%), Financial and insurance services (5.29; 29.4%)
Moderate (3–5 pp)	Postal and telecommunications services (4.86; 42.7%), Other community, social and personal services (4.19; 36.7%), Radio, television, communication and related equipment (3.89; 38.6%), Industrial machinery (3.84; 41.7%), Office and computing machinery and supplies (3.84; 25.9%), Mining, basic metals and related products (3.74; 34.9%), Repair and maintenance services (3.59; 46.1%), Software package and information systems (3.28; 45.6%), Furniture and domestic appliances (3.22; 20.0%), Public utilities (3.12; 49.6%), Leather, textile, plastic and rubber materials (3.12; 27.4%), Laboratory, optical and precision equipment (3.11; 58.3%), Medical equipment and pharmaceuticals (3.03; 47.6%), Food, beverages and tobacco (3.00; 22.7%)
Small (1–3 pp)	Chemical products (2.96; 41.6%), Recreational, cultural and sporting services (2.72; 48.4%), Machinery for mining and construction (2.48; 42.3%), Electrical machinery and equipment (2.13; 30.9%), Transport equipment (2.10; 43.6%), Agricultural machinery (2.00; 42.9%), Construction work (1.98; 15.9%), Education and training services (1.84; 37.9%), IT services (1.67; 49.0%), Musical instruments, sports goods, games and toys (1.23; 31.8%)
Minimal (≤1 pp)	Security, fire-fighting, police and defence equipment (0.97; 46.2%), Research and development services (0.82; 38.5%), Construction materials and auxiliary products (0.52; 28.3%), Hotel, restaurant and retail trade services (0.34; 33.2%), Installation services (0.32; 39.9%), Clothing, footwear and accessories (0.05; 31.9%), Health and social work services (0.00; 45.9%), Business services (0.00; 24.6%), Administration, defence and social security services (0.00; 56.3%), Petroleum products, fuel, electricity and energy (0.00; 42.4%), Agricultural products (0.00; 30.9%), Printed matter and related products (0.00; 44.5%), Transport services (0.00; 32.9%), Oil and gas industry services (0.00; 58.4%), Real estate services (0.00; 34.8%), Supporting and auxiliary transport services (0.00; 33.4%).
Note: Values in parentheses report the median savings difference (percentage points- p.p.) and the share of observed single-offer tenders within the CPV division. Differences were evaluated using Mann–Whitney U tests with Benjamini–Hochberg adjustment; all 45 comparisons remained statistically significant.	

Source: authors

These categories therefore exhibited the strongest separation between observed single-offer and observed multiple-offer procurement outcomes. In contrast, several procurement categories exhibited only minimal differences between the two offer groups. These included Health and Social Work Services, Administration, Defence and Social Security Services, Business Services, Petroleum Products and Energy, and several transport-related categories, where median savings were identical or nearly identical across observed single-offer and observed multiple-offer tenders. Five CPV divisions exhibited median savings differences exceeding five p.p., whereas sixteen divisions exhibited differences of one percentage point or less. These findings indicate that differences between observed single-offer and observed multiple-offer tenders are present across a broad range of procurement markets, although their magnitude varies considerably across procurement categories.

5. Discussion and conclusion

This study examined procurement savings in observed single-offer and observed multiple-offer tenders using TED Contract Award Notices. The results consistently show that observed multiple-offer tenders achieve higher procurement savings than observed single-offer tenders. The difference was statistically significant across all analyses and remained visible in median values, distributional tests, electronic auctions, framework agreements, and all 45 examined CPV divisions. The findings support both auction theory and previous empirical research suggesting that competition strengthens price discipline and leads to more favourable procurement outcomes. Electronic auctions were associated with higher savings, particularly among multiple-offer tenders, although their effect was smaller than the difference associated with the presence of multiple offers. Framework agreements were associated with lower savings, especially among multiple-offer tenders, indicating that their pricing dynamics differ from those of stand-alone procurement procedures. The CPV analysis showed that the relationship between observed competition and procurement savings is widespread but heterogeneous. The largest differences were observed in engineering, water, agricultural, environmental, and financial services, while several administrative, health-related, and transport-related categories exhibited only minimal differences. The supplementary analyses indicate that the relationship between observed competition and procurement savings remains visible across different procurement settings. Although electronic auctions were associated with higher procurement savings in both offer groups, the difference between observed single-offer and observed multiple-offer tenders was substantially larger than the difference associated with auction usage alone. Similarly, procurement savings were lower within framework agreements than outside them, particularly among observed multiple-offer tenders. These findings suggest that procurement savings vary not only with the use of specific procurement mechanisms but also with the level of observed competition. Statistically significant differences between observed single-offer and observed multiple-offer tenders were identified across all 45 examined CPV divisions after adjustment for multiple comparisons. However, the magnitude of these differences varied considerably. The largest median differences were observed in architectural, construction, engineering and inspection services, collected and purified water, agricultural services, environmental services, and financial and insurance services. In contrast, several procurement categories exhibited only minimal differences between the two offer groups. These results indicate that the association between observed competition and procurement savings is present across a broad range of procurement markets, although its magnitude differs substantially across procurement categories.

Several limitations should be acknowledged. The analysis is restricted to tenders with observed offer-count information and therefore does not estimate the overall prevalence of single-bid procurement. Procurement savings represent price movement relative to estimated values and should not be interpreted as a direct measure of efficiency or value for money. In addition, trimming extreme

observations may have excluded a small number of valid cases. Despite these limitations, the study contributes to the procurement literature by shifting attention from the incidence of single-bid procurement toward its economic consequences. The findings indicate that limited observed competition is consistently associated with lower procurement savings, although the magnitude of this relationship varies across procurement markets. Future research could extend the analysis to country-specific settings, contracting-authority characteristics, procurement procedures, and longitudinal designs.

References

Bosio, E., Djankov, S., Glaeser, E., & Shleifer, A. (2022). Public procurement in law and practice. American Economic Review, 112(4), 1091–1117. https://www.nber.org/papers/w27188

Bulow, J., & Klemperer, P. (1996). Auctions versus negotiations. American Economic Review, 86(1), 180–194. https://www.jstor.org/stable/2118262

Carril, R., Gonzalez-Lira, A., & Walker, M. S. (2026). Competition under incomplete contracts and the design of procurement policies. American Economic Review, 116(2), 535–581. https://doi.org/10.1257/aer.20221345

Coviello, D., & Mariniello, M. (2014). Publicity requirements in public procurement: Evidence from a regression discontinuity design. Journal of Public Economics, 109, 76–100. https://10.1016/j.jpubeco.2013.10.008

De Silva, D. G., Dunne, T., Kankanamge, A., & Kosmopoulou, G. (2008). The impact of public information on bidding in highway procurement auctions. European Economic Review, 52(1), 150–181. https:// 10.1016/j.euroecorev.2007.07.003

Delina, R., & Macik, M. (2023). Quality of artificial intelligence driven procurement decision making and transactional data structure. Quality innovation prosperity, 27(1), 103-118. https://10.12776/QIP.V27I1.1819

Dráb, R., Štofa, T., & Delina, R. (2022). Analysis of the efficiency of electronic reverse auction settings: big data evidence. Electronic Commerce Research, 22(2), 427-450. https://10.1007/s10660-020-09433-0

Drake, L., Lundberg, S., & Stentoft, L. (2025). Contractual requirements and bidding behavior in procurement auctions with endogenous entry. Journal of Public Economic Theory, 27, e70019. https://onlinelibrary.wiley.com/doi/10.1111/jpet.70019

Fazekas, M., & Blum, J. R. (2021). Improving public procurement outcomes. Policy research working paper, 2(4), 2-3.

Matilla-García, M., & Vega, M. (2024). An empirical analysis of the impacts of competition on procurement costs. Applied Economics, 57(40), 6258–6273. https://www.tandfonline.com/doi/full/10.1080/00036846.2024.2382934

McAfee, R. P., & McMillan, J. (1987). Auctions and bidding. Journal of Economic Literature, 25(2), 699–738. https://www.jstor.org/stable/2726107

Tátrai, T., Vörösmarty, G., & Juhász, P. (2024). Intensifying competition in public procurement. Public Organization Review, 24(1), 237-257. https://10.1007/s11115-023-00742-0

Thaler, R. H. (1988). Anomalies: The winner's curse. Journal of Economic Perspectives, 2(1), 191–202. https://www.jstor.org.org/stajne/1942752

Titl, V. (2023). The One and Only: Single-Bidding in Public Procurement. USE Working Papers Series, 23(08), 1-45.

Vickrey, W. (1961). Counterspeculation, auctions, and competitive sealed tenders. Journal of Finance, 16(1), 8–37. https://onlinelibrary.wiley.com/doi/10.1111/j.1540-6261.1961.tb02789.x

FROM CAMPUS TO COMPANY: COLLABORATIVE INNOVATION IN THE AGE OF AI

FROM CAMPUS TO COMPANY – COLLABORATIVE INNOVATION IN THE AGE OF AI

Radan Baše, Ondřej Čuka, Ondřej Skřehota

Brighten Digital, s.r.o., Prague
radan.base@brighten.digital, ondrej.cuka@brighten.digital, ondrej@brighten.digital

Miroslav Jaroš, Tomáš Pitner, Oldřich Trenz

Mendel University in Brno
Faculty of Business and Economics
miroslav.jaros@mendelu.cz, tomas.pitner@mendelu.cz, oldrich.trenz@mendelu.cz

DOI: 10.35011/IDIMT-2026-341

Keywords

Artificial intelligence; university-industry collaboration; topic modelling; CRM analytics; multilingual NLP; pharmaceutical sector.

Abstract

The paper presents a practice-oriented case of university-industry collaboration in applied artificial intelligence built around the development of Fieldloop Engine v3.1, a multilingual natural language processing pipeline for the analysis of CRM records in the pharmaceutical sector. The joint objective was to transform fragmented textual notes created by sales representatives into structured thematic intelligence without exposing personal data and without reducing the workflow to simple keyword reporting. The resulting system combines transformer-based embeddings, UMAP dimensionality reduction, HDBSCAN clustering, class-based TF-IDF topic representation, LLM-assisted topic naming, and named entity pseudonymization for Czech, German, and English texts. The engine tracks continuity across chronological releases and detects narrative shifts such as emerging themes, keyword drift, and volume changes. The paper argues that the technical architecture and the collaboration architecture were equally important to the project outcome. The case shows that collaboration creates the greatest value when it links methodological rigor, domain-specific constraints, privacy requirements, and deployable engineering in a shared iterative pipeline.

1. Introduction

Artificial intelligence adoption is increasing pressure on organisations to turn unstructured textual data into operational knowledge (Zaoui Seghroucheni et al., 2025). This pressure is especially visible in CRM environments, where employees record observations, objections, and contextual signals in short free-text notes (Galitsky, 2020; Ozan, 2021). Such records often contain strategically valuable information, but they remain difficult to analyze at scale because manual coding is expensive, predefined reporting categories are too rigid, and generic large language model workflows may not

fit multilingual and privacy-sensitive data. Recent work therefore increasingly explores NLP-based approaches to CRM intelligence (Sabapathy, 2025; Sundstedt, 2024).

The joint solution, documented in the project repository as *Fieldloop Engine*, was designed to discover recurring topics in Czech, English, and German text, attach interpretable labels, pseudonymize named entities, and track how themes change across successive data releases. The business need was clear: *transform dispersed mention-level notes into a form that supports trend detection and decision-making*. The academic contribution was equally clear: bring together *current topic-modelling methods, privacy-aware preprocessing, and evaluation logic in a replicable pipeline.*

2. Collaboration context and problem definition

2.1. Business problem

The collaboration addressed a familiar but under-analyzed issue in CRM practice. Pharmaceutical sales teams produce large numbers of short textual mentions after meetings and follow-up interactions. Individually, these notes are locally useful; collectively, they form a longitudinal narrative of the market. The difficulty is that the data are noisy, multilingual, uneven in detail, and entangled with personal information (Karmalkar et al., 2021). From an industry perspective, the need was not merely to classify notes into a fixed taxonomy, but to discover emerging themes, track continuity across releases, and communicate changes in a form interpretable to non-technical stakeholders.

2.2. Research and design problem

From an academic perspective, the problem combined multilingual processing, short-text topic discovery, privacy-preserving preprocessing, temporal continuity, and human-readable outputs. The solution was not intended as a benchmark prototype. It had to operate on realistic JSON-based CRM inputs, process chronological releases in batch mode, preserve lineage between topics, and fit into a delivery model that reduced operational friction. Given the high-stakes nature of the pharmaceutical industry, strategic and operational decision-making demands robust, reliable, and trustworthy analytical support.

2.3. Collaborative alignment

This made the case particularly suitable for campus-to-company collaboration. The company side contributed a concrete domain problem, input structure, expected outputs, and operational constraints. The university side contributed current NLP methods and an analytical framework for balancing accuracy, interpretability, and privacy. In implementation terms, the solution rests on three pillars: the *NLP Engine* for discovery and traceability, a *Synthetic Data Simulator* for controlled validation before exposure to live CRM data, and the *MediaHouse* publication layer for business-facing delivery. The case therefore illustrates a broader pattern: useful AI innovation emerges when method and application are translated jointly rather than handed off from one side to the other.

2.4. Methodological approach

This paper follows an exploratory case-study approach. The analysis is based on the design, implementation, and deployment of the Fieldloop Engine developed jointly by Brighten Digital and Mendel University. Technical documentation, architectural artifacts, implementation repositories, and deployment experiences were used as primary sources for reflection and analysis.

3. Solution architecture and technology stack

Fieldloop Engine's core logic follows a *BERTopic-style* pattern: embedding, dimensionality reduction, density-based clustering, and class-based TF-IDF topic representation (Grootendorst, 2022).

3.1. Processing stages

The engine processes data in six stages:

1. Input loading, validation, entity anonymization, and boilerplate detection.
2. Document embedding, dimensionality reduction through UMAP, clustering through HDBSCAN, and keyword extraction through class-based TF-IDF (McInnes et al., 2018; McInnes et al., 2017). Stage 1b enriches the current release with historical context by retrieving related earlier topics from embedded Qdrant storage and calculating lineage.
3. LLM-based assignment of intelligible labels, descriptions, and business dimensions.
4. Optional evaluation against ground truth using quality measures such as NMI, ARI, and silhouette values.
5. Narrative-shift detection, including topic emergence, volume increases, volume decreases, and keyword drift.
6. Writing the resulting structured outputs, including topics.json, narrative_shifts.json, mention_classifications.json, and run_metadata.json.

3.2. Traceability and data model

This staged design combines research logic with operational needs. Rather than stopping at clustering, the implemented pipeline preserves analytical traceability: it makes visible where *privacy* measures are applied, how *topics* are formed, and how *historical continuity* influences interpretation. The input model is also significant. Each JSON file represents one mention and may include text, timestamps, account metadata, contact metadata, salesperson metadata, and business weights. The system therefore analyses language together with light business context instead of treating all notes as isolated text fragments.

3.3. Technology stack

The stack is deliberately heterogeneous. It combines a Python-based NLP pipeline, Docker-based execution, provider-based model access, multilingual NER sidecars, and embedded Qdrant storage. Analytically, the system follows a BERTopic-style workflow built from embeddings, UMAP, HDBSCAN, class-based TF-IDF, LLM labeling, and lineage tracking (Grootendorst, 2022; McInnes et al., 2018; McInnes et al., 2017). Operationally, the key design choice is modularity: model providers can be exchanged, batch processing remains reproducible, and historical state is preserved across releases instead of being lost after a single run.

3.4. Reporting and delivery layer

The business-facing output is not presented primarily as a dashboard, but as an editorial publication layer called *MediaHouse*. The weekly output combines a masthead, hero story, thematic sections, editorial commentary, contributor bylines, and digest-style summaries. This matters because the

delivery format performs part of the interpretive work for the reader: it ranks signals, foregrounds one narrative as decisive, and turns clusters into a bounded account of the market. In cognitive terms, it favors narrative processing and aligns with the higher-level construal of weekly strategic interpretation rather than real-time operational monitoring (Bruner, 1986; Kahneman, 2011; Trope & Liberman, 2010).

That same strength creates risk. A polished editorial output is easier to trust and quicker to act on than a confusing analytical screen. If a weakly supported topic is elevated into a prominent narrative slot, the resulting managerial response can be faster and more consequential than in a slower interface. The delivery layer therefore sharpens the need for upstream rigor rather than relaxing it.

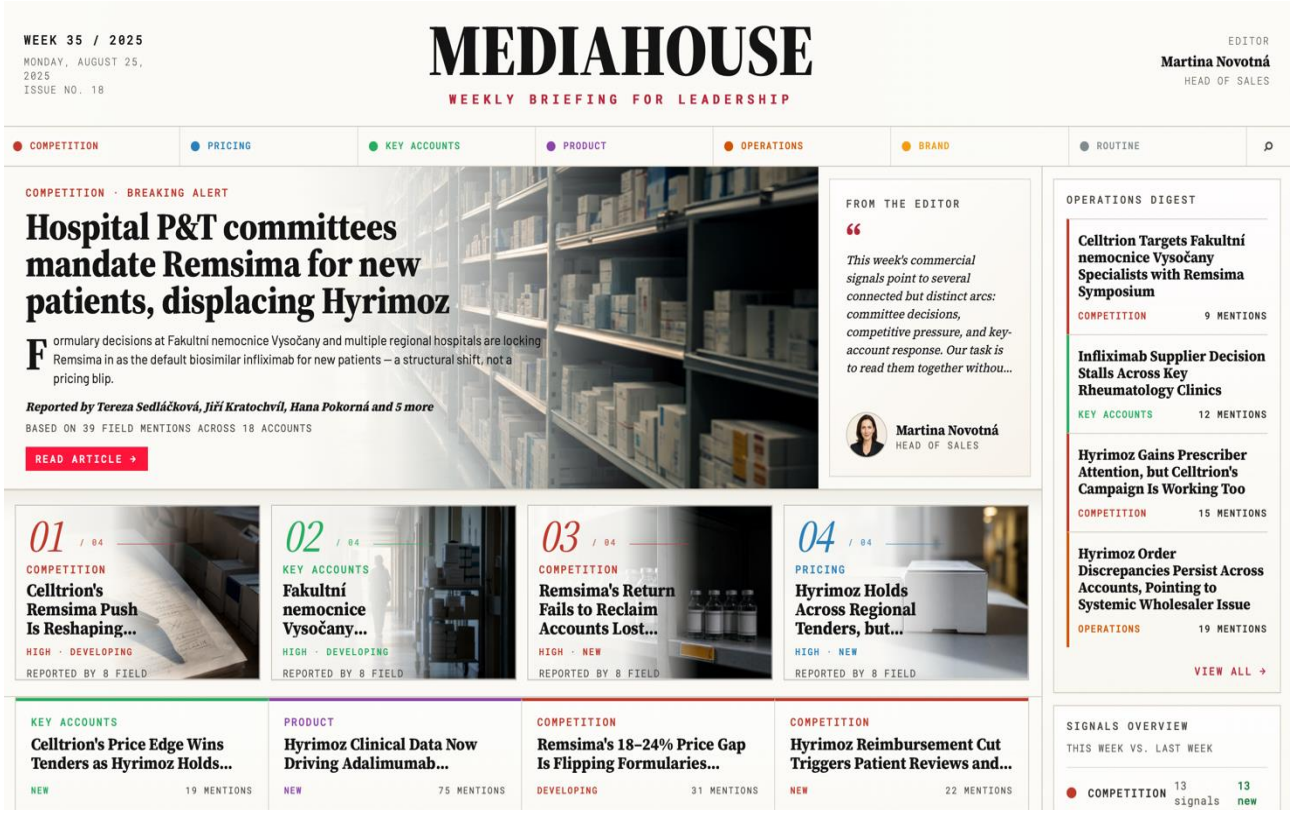


Figure 1. MediaHouse publication layer: weekly editorial delivery format for strategic sales intelligence
Source: authors, screenshot from the MediaHouse publication layer

4. Privacy, multilingualism, and interpretability

4.1. Privacy by design

One of the strongest aspects of the case is that privacy and interpretability were treated as design requirements rather than post hoc concerns. Before embeddings are generated, the engine removes overly short texts, flags boilerplate CRM templates, strips titles, and pseudonymizes named entities. Named entity recognition is handled through a Docker sidecar service combining a Slavic-language model and an English model, with regex fallback when necessary. The pseudonymization step replaces personal, organizational, and location references with placeholders such as [PERSON], [ORG], and [LOC].

Figure 2 highlights the operational architecture behind these controls. Privacy is not an isolated preprocessing trick. It is connected to the NER sidecar, the `text_clean` representation stored in the vector layer, the de-pseudonymization step applied only to final outputs, and the reporting layer. This co-design is central to the argument of the paper.

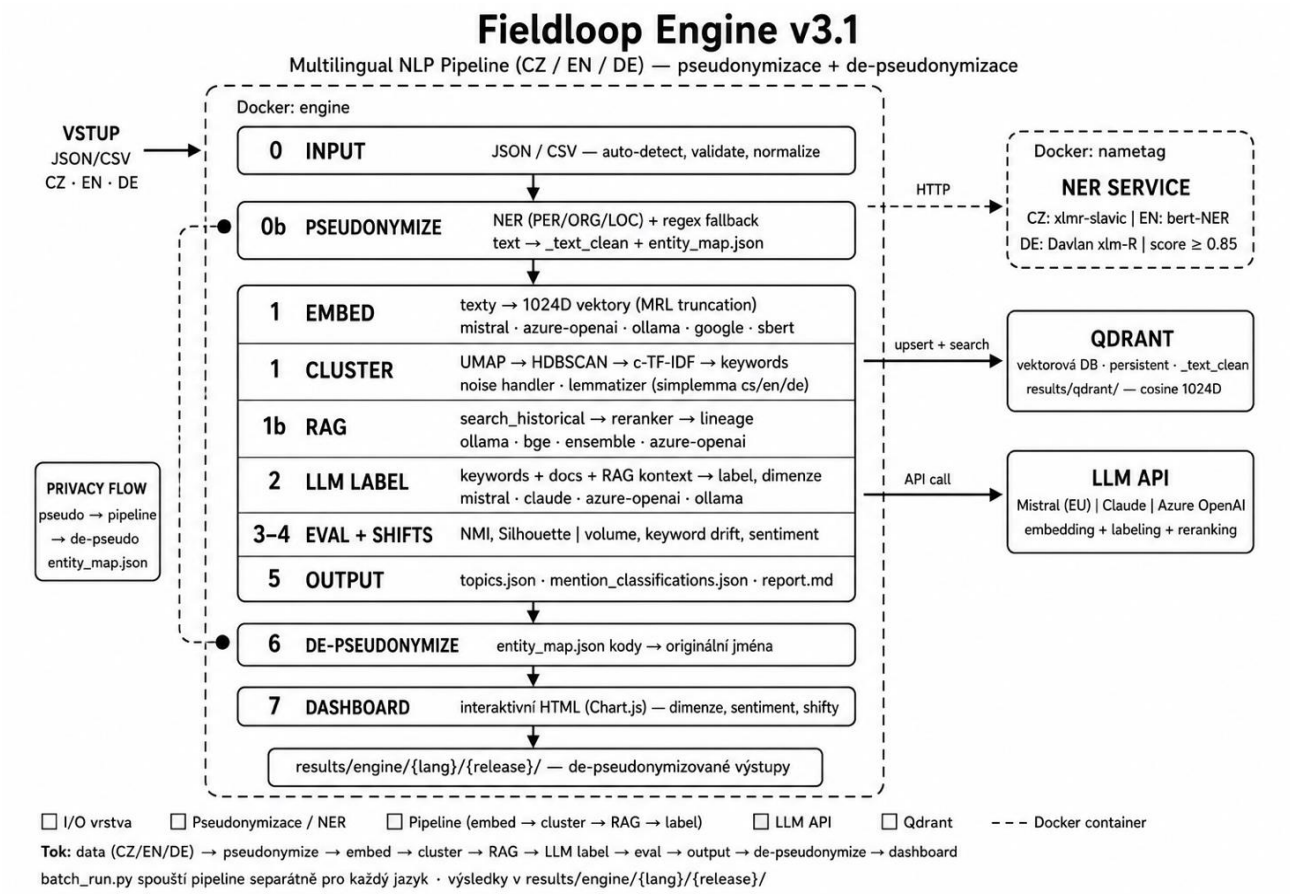


Figure 2. Privacy-preserving and operational integration detail derived from the repository architecture diagram
Source: authors, based on the project repository architecture diagram

In healthcare-adjacent commercial environments, free-text CRM data can contain names of physicians, institutions, and location cues. By applying pseudonymization before embedding and keyword extraction while preserving original text for outputs and retrieval logic, the pipeline balances analytical usefulness with privacy protection.

4.2. Multilingual processing

Multilingualism presents a second challenge. The engine is intended for Czech, English, and German inputs. Designing such a system means more than selecting a broadly multilingual model. It requires preprocessing, labeling, and evaluation that preserve comparability across languages without flattening local nuances.

4.3. Interpretability for adoption

Interpretability forms the third design pillar. The engine does not stop at cluster formation. It produces human-readable labels, descriptions, business dimensions, keywords, and continuity indicators that can be discussed in business language. In the deployed product, interpretability extends into the

editorial form of the output itself. The weekly publication format provides narrative synthesis instead of exploratory analytics, which supports fast strategic reading but also creates a stronger obligation for upstream rigor. Weak clustering, poor lineage logic, or inadequate privacy controls would be amplified into persuasive but potentially misleading managerial narratives. This risk is consistent with research on heuristic judgement, narrative fallacy, tightly coupled decision systems, and normalization of deviance under speed pressure (Tversky & Kahneman, 1974; Taleb, 2007; Perrow, 1984; Vaughan, 1996). The engineering rigor documented in this paper is therefore not only enabling infrastructure beneath reporting. It is also protective infrastructure that makes high-velocity interpretive delivery ethically and commercially defensible.

4.4. Continuity and narrative-shift detection

A distinctive feature of the architecture is its treatment of time. Instead of analyzing each release independently, the batch runner processes releases chronologically and accumulates state in embedded Qdrant storage. Each new release can therefore be interpreted against the lineage of previous releases. The output schema captures this orientation through continuity scores, trends, and narrative-shift categories such as volume increase, volume decrease, keyword drift, and topic emergence.

5. Collaborative innovation in the age of AI

The README makes clear that the project was engineered for practical execution, not only demonstration. Docker-based batch execution, mounted configuration, structured outputs, embedded vector storage, and explicit module boundaries make the artifact easier to reproduce, inspect, and hand over. This is relevant to the collaboration argument because many university-industry projects succeed analytically but fail at operational packaging. Here, deployment-oriented engineering is part of the collaborative outcome rather than an afterthought.

5.1. Problem-Centered design

The Fieldloop Engine case highlights several broader lessons about university-industry collaboration.

First, the highest-value AI collaborations start from a sharply defined operational problem, not from generic enthusiasm for the technology. The project was anchored in a clear use case: *multilingual CRM text in a privacy-sensitive pharmaceutical setting*.

Method Integration. Second, contemporary collaboration requires method integration rather than method substitution. In this case, that integration spans the three-pillar architecture rather than the NLP pipeline alone: discovery and lineage in the engine, controlled validation in the synthetic-data layer, and managerial interpretability in the publication layer. The engine combines embeddings, UMAP, HDBSCAN, class-based TF-IDF, LLM labeling, and NER-based pseudonymization because the business problem required all of them (Grootendorst, 2022; McInnes et al., 2018; McInnes et al., 2017).

Governance and Trust. Third, privacy and governance are part of innovation capacity. In this case they shaped the architecture productively through pseudonymization, service isolation, and explicit output schemas.

Longitudinal Intelligence. Fourth, time awareness matters. The lineage and narrative-shift components show that practical AI systems often create value by modelling change rather than only classification.

Shared Readability. Fifth, collaboration succeeds when the artifact can be understood by both partners in their own language. For academic participants, the engine can be described through clustering, evaluation, and topic representation. For business participants, it can be described through themes, alerts, continuity, and reports.

Protective Depth. Sixth, research depth can become an ethical and commercial precondition of polished delivery rather than merely a technical contribution behind it. In domains where deployed outputs accelerate high-stakes decision-making, upstream research and engineering determine whether a fast and persuasive delivery format is defensible. In the present case, privacy-preserving preprocessing, validated clustering, traceable lineage, and narrative-shift detection reduce the risk that a compelling editorial artifact will industrialize confidently wrong decisions at scale.

6. Conclusion

This paper reframed the keynote topic, *From Campus to Company: Collaborative Innovation in the Age of AI*, through a concrete case: the development of Fieldloop Engine for multilingual analysis piloted on pharmaceutical CRM data which paved the road to other industries through complexity and high-stake position of the pharma/healthcare sectors. The case shows that AI collaboration becomes meaningful when a university-industry partnership produces more than a conceptual discussion or a one-off prototype. It must deliver an artifact that encodes methodological rigor, domain knowledge, privacy discipline, temporal reasoning, and operational usability in one integrated design.

The Fieldloop Engine demonstrates such an integration. The *MediaHouse* publication layer makes this point sharper: once AI output is delivered in a polished editorial form that accelerates strategic reading and action, research rigor is no longer just a source of better analysis. It becomes part of the governance structure that protects from fast, persuasive, and costly error.

In the age of AI, the journey from campus to company should not be understood as a one-directional transfer of knowledge. It is a co-creation process in which research methods are reshaped by deployment realities and business needs are clarified through analytical discipline (Cohen & Levinthal, 1990; Etzkowitz & Leydesdorff, 2000). The more AI enters sensitive and multilingual organizational environments, the more valuable this kind of collaboration becomes. The strongest innovation partnerships will be those that can turn that shared complexity into maintainable, interpretable, and trustworthy systems, and that recognize a further point made visible by this case: in high-stakes environments, polished delivery is only as legitimate as the protective depth of the research and engineering beneath it.

Acknowledgement

This work was supported by the project Fieldloop.ai: Development of Intelligent Automation for CRM Systems (Fieldloop.ai: Vývoj inteligentní automatizace pro CRM systémy), project No. CZ.01.01.01/01/24_062/0007568, funded under the Operational Programme Technologies and Applications for Competitiveness, administered by the Ministry of Industry and Trade of the Czech Republic and co-financed by the European Union. The authors acknowledge the collaboration and support between Brighten Digital (https://www.brighten.digital) and Mendel University (https://informatika.mendelu.cz).

References

Cohen, W. M., & Levinthal, D. A. (1990). Absorptive capacity: A new perspective on learning and innovation. Administrative Science Quarterly, 35(1), 128-152. https://doi.org/10.2307/2393553

Bruner, J. (1986). Actual minds, possible worlds. Harvard University Press.

Etzkowitz, H., & Leydesdorff, L. (2000). The dynamics of innovation: From national systems and "Mode 2" to a triple helix of university-industry-government relations. Research Policy, 29(2), 109-123. https://doi.org/10.1016/S0048-7333(99)00055-4

Galitsky, B. (2020). Artificial intelligence for customer relationship management: Keeping customers informed. Springer. https://doi.org/10.1007/978-3-030-52167-7

Grootendorst, M. (2022). BERTopic: Neural topic modeling with a class-based TF-IDF procedure. arXiv. https://arxiv.org/abs/2203.05794

Kahneman, D. (2011). Thinking, fast and slow. Farrar, Straus and Giroux.

Karmalkar, P., Gurulingappa, H., Muhith, J., Singhal, S., Megaro, G., & Buchholz, F. (2021). Improving consumer experience for medical information using text analytics. In 2021 International Symposium on Electrical, Electronics and Information Engineering (pp. 471-476).

McInnes, L., Healy, J., & Astels, S. (2017). hdbscan: Hierarchical density-based clustering. Journal of Open-Source Software, 2(11), 205. https://doi.org/10.21105/joss.00205

McInnes, L., Healy, J., Saul, N., & Grossberger, L. (2018). UMAP: Uniform manifold approximation and projection. Journal of Open-Source Software, 3(29), 861. https://doi.org/10.21105/joss.00861

Ozan, S. (2021). Case studies on using natural language processing techniques in customer relationship management software. Journal of Intelligent IS, 56(2), 233-253. https://doi.org/10.1007/s10844-020-00619-4

Perrow, C. (1984). Normal accidents: Living with high-risk technologies. Princeton University Press.

Sabapathy, G. (2025). Artificial Intelligence in Pharmaceutical Commercial Operations: Transforming Sales Performance and Patient Engagement in the Digital Era. Journal of Computer Science and Technology Studies, 7(8), 636-651.

Sundstedt, V. (2024). Large language models for data-driven customer relationship management: Transforming unstructured data into business intelligence [master's thesis]. Doria. https://urn.fi/URN:NBN:fi-fe2024091170617

Taleb, N. N. (2007). The Black Swan: The impact of the highly improbable. Random House.

Trope, Y., & Liberman, N. (2010). Construal-level theory of psychological distance. Psychological Review, 117(2), 440-463. https://doi.org/10.1037/a0018963

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. Science, 185(4157), 1124-1131. https://doi.org/10.1126/science.185.4157.1124

Vaughan, D. (1996). The Challenger launch decision: Risky technology, culture, and deviance at NASA. University of Chicago Press.

Zaoui Seghroucheni, O., Lazaar, M., & Al Achhab, M. (2025). Using AI and NLP for tacit knowledge conversion in knowledge management systems: A comparative analysis. Technologies, 13(2), 87. https://doi.org/10.3390/technologies13020087

BUSINESS, TRADE AND MARKETING IN THE AGE OF AI

AI IN SUSTAINABLE BRANDING: CONSUMER BEHAVIOUR IN THE ORGANIC PRODUCTS MARKET

Mária Janošková, Erik Weiss, Jozef Zuzik

Bratislava University of Economics and Business, Slovakia
Faculty of Business Economics with seat in Kosice
maria.ria.janoskova@euba.sk, erik.weiss@euba.sk, jozef.zuzik@euba.sk

DOI: 10.35011/IDIMT-2026-351

Keywords

Artificial Intelligence; Consumer behaviour; Organic products market; Competitive advantage; Motives; Values.

Abstract

Consumer behaviour in the organic products market is shaped by a combination of rational and behavioural factors, with health, trust, and environmental values playing a key role. In the context of digital transformation, however, artificial intelligence is becoming increasingly involved in this process, changing the way brands communicate, distribute products, and build relationships with customers. The aim of this paper is to analyse the determinants of consumer behaviour in the organic products market and to identify the possibilities of linking them with the use of artificial intelligence in marketing practice. The research is based on a combination of secondary analysis of scientific sources and a quantitative survey conducted in Slovakia. The results indicate the dominant importance of trust, perceived quality, and health benefits, while price remains the main barrier to purchase. At the same time, it appears that artificial intelligence tools have the potential to increase the effectiveness of marketing communication, content personalisation, and product availability; however, their use may influence the perceived authenticity of a brand. The paper highlights the need for a balanced approach to the implementation of AI in the organic segment, where competitive advantage is not generated solely by technological advancement, but also by the ability to preserve trust and transparency in relation to the consumer.

1. Introduction

The organic products market currently represents an important segment of the food market, the development of which is determined not only by increasing living standards but also by changes in value orientations and attitudes, with an emphasis on health, quality, and environmental sustainability. Consumer behaviour in this segment is characterised by a specific complexity, as purchasing decisions are influenced by a combination of rational considerations regarding price and availability, value-based preferences, and psychological aspects. Despite the growing interest in organic products, however, a discrepancy persists between positive attitudes and actual purchasing behaviour, which points to the need for a deeper understanding of the mechanisms underlying consumer decision-making.

In the context of digital transformation, this issue acquires a new dimension. Artificial intelligence is fundamentally changing the way companies analyse consumer behaviour, personalise marketing communication, and optimise distribution processes. In the organic products segment, where trust, transparency, and authenticity are crucial, the implementation of AI represents not only an opportunity but also a challenge. On the one hand, it enables more effective targeting, the reduction of inefficiencies in marketing activities, and the support of sustainable marketing strategies through resource optimisation. On the other hand, it may influence brand perception, particularly if communication is perceived as overly automated or inauthentic.

From this perspective, the competitive advantage of companies in the age of artificial intelligence is not derived solely from technological maturity in the use of data, but from the ability to integrate AI into marketing strategies and processes in a way that strengthens trust, respects consumers' value-oriented decision-making, and at the same time contributes to the sustainability of marketing activities. Therefore, the aim of this paper is to analyse the determinants of consumer behaviour in the organic products market and to identify their connection with the use of artificial intelligence, with an emphasis on the development of sustainable marketing strategies and the building of competitive advantage in the digital environment.

2. Literature review

Research on consumer behaviour in the organic products market is characterised by a high degree of interdisciplinarity, integrating insights from marketing, behavioural economics, and environmental studies (Sheth et al., 1991). Consumer decision-making in this segment cannot be explained solely through rational economic models; rather, it is the result of the interaction of cognitive, functional, social, and emotional values, as well as situational influences (Rana & Paul, 2017). From the perspective of consumer motivation, the literature distinguishes between egoistic and altruistic motives. Egoistic motives are mainly associated with health, food safety, and quality of life, whereas altruistic motives include environmental protection, ethical principles, and support for sustainable production (Hughner et al., 2007). Food consumption often extends beyond a utilitarian framework and becomes an expression of identity, lifestyle, and the value orientation of the individual.

Empirical findings have long confirmed that although environmental aspects play an important role in the context of growing societal sensitivity to sustainability issues (Hughner et al., 2007), health benefits dominate in many cases (Magnusson et al., 2003). Consumers perceive organic food as safer, of higher quality, and more beneficial to health. Rana and Paul (2017) systematise the determinants of purchasing behaviour and identify health as the dominant motivational factor, while increasing health risks and an orientation towards a healthy lifestyle strengthen preferences for organic products. Yadav and Pathak (2016) examined young consumers' intentions to purchase ecological products. Their research findings confirmed that health consciousness, together with environmental attitudes, plays a significant role. Trust is a key factor in transforming interest into actual purchase. It is primarily linked to certification, product origin, quality, and the transparency of the supply chain. Credible labelling and clear communication significantly support consumers' willingness to purchase organic products. Conversely, a lack of trust may weaken the relationship between a positive attitude and purchasing behaviour.

The main barriers to the wider adoption of organic products include higher price, limited availability, and information asymmetry. Gleim et al. (2013) argue that even when consumers hold positive attitudes, practical barriers may significantly reduce actual purchasing. Price remains one of the most important obstacles to the growth of the organic products market, while consumers' willingness to

pay a price premium differs according to income, values, and market segment (Aschemann-Witzel & Zielke, 2017; Aertsens et al., 2011).

An important topic in current research is the so-called attitude–behaviour gap, that is, the discrepancy between declared positive attitudes towards organic products and actual purchasing behaviour. Chekima et al. (2023) refer to the "green gap", whereby consumers declare support for sustainable consumption but, in practice, make decisions under the influence of price, convenience, or distrust. Similar conclusions are presented by Carrington et al. (2014), who draw attention to the difference between a positive attitude towards organic products and actual purchasing decisions. Zhuo et al. (2022) and Wintschnig (2021) also emphasise the importance of moral and ethical aspects of consumer decision-making. Although consumers often declare a positive attitude towards organic products, their actual purchasing behaviour is influenced by a broader range of factors that go beyond the framework of rational decision-making (Csikosova et al., 2018).

Digital transformation is fundamentally changing the way consumers search for information, compare products, and communicate with brands. The digital environment redefines customer experience through omnichannel interactions, personalisation, and data analytics (Lemon & Verhoef, 2016). In the organic segment, digitalisation increases product availability, facilitates comparison of offers, and enables more transparent communication about the origin, quality, and environmental attributes of products. In this context, artificial intelligence becomes increasingly important, as it changes marketing communication, content strategies, and customer relationship management. Huang and Rust (2021) point to the growing importance of AI in personalisation, prediction of consumer preferences, and automation of marketing processes. Davenport et al. (2020) emphasise that AI can increase marketing efficiency, while at the same time raising questions related to privacy protection, algorithmic bias, and ethics. A particularly sensitive issue in the organic segment is trust in relation to AI. Longoni et al. (2019) found that consumers distrust artificial intelligence and prefer human decision-making over algorithmic decision-making, especially in areas connected with health and empathy. This is also relevant to organic products, where health and authenticity are key values (Paschen et al., 2020). Excessive automation of marketing communication may therefore weaken the perceived authenticity of the brand and emotional bonds with the customer. This suggests that the use of AI may paradoxically weaken the perceived authenticity of the brand.

Despite international research, there remains a lack of empirical studies focused on smaller Central European markets, including Slovakia. Consumer behaviour in these markets may be influenced by different levels of purchasing power, price sensitivity, trust in certification, or digital adoption. The research gap lies in the need to analyse the determinants of organic product purchases in the local environment and, at the same time, to assess the extent to which artificial intelligence tools can support competitive and sustainable marketing strategies without undermining trust and transparency in relation to the consumer.

3. Methodology

The aim of the research is to analyse the determinants of consumer behaviour in the organic products market and to identify the possibilities of linking them with the use of artificial intelligence in marketing practice. The methodological approach is based on a combination of secondary and primary research, thereby ensuring a more comprehensive understanding of the examined issue and the triangulation of findings.

The research is designed as an exploratory-descriptive study with elements of comparative analysis. The selected design makes it possible to identify the main determinants of organic product purchases, assess the significance of individual purchase barriers, analyse consumers' attitudes towards the use

of AI in marketing, and propose implications for sustainable marketing strategies. The research was conducted in two successive stages.

3.1. Secondary data analysis

The first stage consists of a systematic literature review and a secondary analysis of available sources. The analysis focused on:

- Scientific articles indexed in Web of Science and Scopus,
- Information and data on the organic food market,
- Reports by professional and research institutions,
- Studies on the use of AI in trade and marketing.

The secondary analysis focused on identifying recurring determinants of consumer behaviour, particularly in the areas of health, trust, environmental values, price, availability, digitalisation of the purchasing process, trust in AI, and algorithmic transparency. The outputs of the secondary analysis served as the theoretical basis for formulating the research questions and constructing the questionnaire.

3.2. Primary quantitative research

The second stage of the research consists of a questionnaire survey conducted in Slovakia. The research sample comprised 153 respondents, representing consumers with varying levels of experience in purchasing organic products. Data collection was carried out electronically through a structured Google Forms questionnaire. The questionnaire included closed-ended and scale-based questions focused on:

- Identifying the determinants and frequency of organic product purchases,
- Purchase motives,
- Purchase barriers,
- The level of trust in organic labelling, certification, and product transparency,
- Attitudes towards personalised digital recommendations,
- Perceptions of the use of artificial intelligence in consumer behaviour.

For value-based items, a Likert scale was used (1–5), where 1 indicated complete disagreement and 5 indicated complete agreement.

3.3. Research questions

Based on the secondary analysis, the following research questions were formulated:

- RQ1: Which factors most significantly influence the purchase of organic products in the Slovak market?
- RQ2: Does price represent the main barrier to purchasing organic products?
- RQ3: What is the significance of trust in certification and product transparency?
- RQ4: How do consumers perceive the use of artificial intelligence in the marketing of organic products?
- RQ5: Can AI contribute to the competitive advantage of companies in the organic segment without weakening consumer trust?

The selected methodology enables the connection of traditional determinants of consumer behaviour with new digital factors represented by artificial intelligence. It thus provides a basis for identifying modern sustainable marketing strategies and competitive advantages of companies operating in the organic products market.

4. Results

The research sample consisted of 206 respondents. The results indicate that the collected data primarily reflect the views of younger age groups, which should be taken into account when interpreting the findings, particularly in relation to attitudes towards digitalisation and the use of artificial intelligence in marketing. In terms of age structure, the dominant group comprised respondents under the age of 30, representing 81.1% of the research sample (167 respondents). This is predominantly a younger generation of consumers, characterised by a higher level of digital literacy, openness to innovation, and a growing interest in a healthy lifestyle. The majority of respondents were women, numbering 146, i.e. 70.9%, while men accounted for 60 respondents, i.e. 29.1%. According to place of residence, 47.6% of respondents lived in rural areas and 52.4% in urban areas.

Table 1. Sociodemographic characteristics of the research sample

Age category	Number	Share (%)
Up to 20 years	9	4.4
21-30 years	158	76.7
31-40 years	22	10.7
41-50 years	11	5.3
51-60 years	6	2.9
61 years and over	-	-
Total	206	100 %

Source: authors

4.1. Evaluation of the research questions

Based on the questionnaire survey conducted in the Slovak market (n = 206) and in accordance with the aim of the paper, the research questions were evaluated as follows:

RQ1: *Which factors most significantly influence the purchase of organic products in the Slovak market?*

- The results showed that the most significant factors influencing the purchase of organic products are health- and quality-related motives. As many as 88.8% of respondents stated that they choose organic products because they consider them healthier, of higher quality, and free from pesticides. This result confirms the dominance of rational and utilitarian purchase motives.
- The second significant factor was environmental values, with 47.5% of respondents indicating environmental friendliness as a reason for purchase.
- Another factor was the perception of higher freshness and better taste (28.7%).

The findings suggest that Slovak consumers combine health expectations, perceived quality, ecological attitudes, and the sensory benefits of the product when purchasing organic products.

Conclusion to RQ1: The purchase of organic products is primarily influenced by health, quality, and trust in the product, and secondarily by environmental values.

RQ2: *Does price represent the main barrier to purchasing organic products?*

- The results clearly confirmed that price is the dominant barrier to purchasing organic products. As many as 78.8% of respondents identified higher prices as the main reason preventing them from purchasing or making their decision more difficult.

High price sensitivity indicates that even when consumers have a positive attitude towards organic products, their price may hinder actual demand. This represents a classic manifestation of the attitude–behaviour gap, i.e. the discrepancy between positive attitudes and actual behaviour.

Conclusion to RQ2: Yes, price represents the main barrier to purchasing organic products in the Slovak market.

RQ3: *What is the significance of trust in certification and product transparency?*

- The results showed that trust represents a key market determinant. As many as 57.5% of respondents know only some organic labels, while 20% do not know any. At the same time, a considerable share of respondents (16%) stated that they are unable to reliably distinguish which products are genuinely organic.
- Another important signal is the preference for content such as the story of the producer and the farm (as much as 65%), as well as transparent information about the origin of the product.

Consumers therefore seek not only a product, but also a credible story, origin, and authenticity.

Conclusion to RQ3: Trust in certification, clear labelling, and transparent communication has strategic importance for the development of the organic products market.

RQ4: *How do consumers perceive the use of artificial intelligence in the marketing of organic products?*

- Consumers' attitudes are derived from their preferences regarding forms of communication and information channels. Respondents prefer personalised and valuable content (56.3%), digital channels, social media, and YouTube (47.5%), web articles and newsletters (41.3%), and authentic brand stories (65%).

These results suggest that consumers are open to digital and AI-supported forms of communication if they provide value, education, and relevance. Conversely, the lower preference for traditional advertising (21.3%) indicates that conventional mass campaigns are losing effectiveness.

Conclusion to RQ4: Consumers perceive AI positively when it serves to personalise communication, improve access to information, recommend relevant products, and simplify decision-making. They perceive AI negatively if it appears inauthentic or manipulative.

RQ5: *Can AI contribute to the competitive advantage of companies in the organic segment without weakening consumer trust?*

- The research findings suggest that it can, but only under certain conditions. AI can create competitive advantage through personalised campaigns, intelligent content marketing, recommendation systems, distribution optimisation, demand prediction, and the clear explanation of product composition.
- At the same time, consumers in the organic segment strongly prefer authenticity, brand storytelling, transparency, and trust. If AI-driven communication appeared artificial, untruthful, or misleading, it could weaken trust.

Conclusion to RQ5: AI can be a source of competitive advantage if it is implemented ethically, transparently, and in accordance with the values of the organic segment.

The primary research confirmed that the Slovak organic products market has growth potential; however, its development is limited by higher prices, information ambiguity, weaker knowledge of certification, and insufficient availability. At the same time, it becomes evident that the future competitive advantage of companies will not be based solely on the product itself, but on the ability to connect trust, transparency, personalisation, AI, and sustainability.

5. Discussion

The obtained results confirm the conclusions of international studies, according to which consumer behaviour in the organic products market is determined by a combination of rational, value-based, and situational factors. The dominance of health benefits, product quality, and product safety corresponds with the findings of Magnusson et al. (2003), Hughner et al. (2007), and Rana and Paul (2017), who identified health as the main motive for purchasing organic food. Thus, even in the Slovak context, environmental aspects appear to be important, although they are often secondary in relation to the consumer's personal benefit.

The pronounced price sensitivity of respondents confirms the persistent issue of the so-called attitude–behaviour gap, that is, the discrepancy between a positive attitude towards organic products and actual purchasing behaviour (Carrington et al., 2014). Although consumers perceive organic products positively, higher price remains a significant barrier to broader market adoption. At the same time, the results emphasised the importance of trust, transparency, and clear product labelling. Lower awareness of certification and uncertainty in identifying genuinely organic products confirm the findings of Janssen and Hamm (2012), according to which trust is an essential prerequisite for the development of the organic market.

In the context of digitalisation, it was shown that consumers prefer authentic and valuable content, producer stories, and online information sources. This creates scope for the use of artificial intelligence in communication personalisation, content marketing, and distribution. At the same time, however, the results indicate that, in the organic segment, the implementation of AI must be sensitive, transparent, and ethical in order not to undermine brand authenticity and consumer trust. Competitive advantage in the age of AI is therefore not created solely through technological advancement, but through the ability to connect innovation with the values of sustainability and trust.

6. Conclusion

In this paper, the authors focused on analysing consumer behaviour in the organic products market and identifying the possibilities of using artificial intelligence in marketing practice. The results confirmed that consumer decision-making is determined primarily by health benefits, perceived quality, trust in the product, and the transparency of information, while price represents the most significant barrier to purchase. At the same time, it was shown that consumers prefer authentic communication, producer stories, and digital sources of information.

In the context of increasing digitalisation, artificial intelligence can contribute to greater efficiency in marketing activities through content personalisation, more precise communication targeting, distribution optimisation, and support for customer decision-making. In the organic segment, however, competitive advantage is not created solely through technological advancement, but through the ability to connect AI-based solutions with trust, transparency, and the values of sustainability.

In conducting the primary research, certain limitations were identified, particularly in relation to the size and structure of the sample, in which younger respondents and women predominated. The results are also linked to the Slovak market environment; therefore, they cannot be immediately generalised to other countries. Future research should focus on international comparisons of consumer behaviour, customer segmentation according to socio-demographic characteristics, and experimental verification of the effects of AI tools on trust, loyalty, and purchasing decision-making. A promising research

topic is also the examination of the impact of generative AI on brand authenticity and consumers' willingness to pay a higher price for organic products.

Acknowledgement

This research was funded by Slovak agency KEGA grant no. 032EU-4/2024 and KEGA grant no. 020EU-4/2025.

References

Aertsens, J., Verbeke, W., Mondelaers, K., & Van Huylenbroeck, G. (2011). The influence of subjective and objective knowledge on attitude, motivations and consumption of organic food. British Food Journal, 113(11), 1353–1378.

Aschemann-Witzel, J., & Zielke, S. (2017). Can't buy me green? A review of consumer perceptions of and behavior toward the price of organic food. Journal of Consumer Affairs, 51(1), 211–251.

Carrington, M.J., Neville. B.A., & Whitwell, G.J. (2014). Lost in translation: Exploring the ethical consumer intention–behavior gap. Journal of Business Research, 67(1), 2759-2767.

Csikosova, A., Janoskova, M., Culkova, K. (2018). Increasing of Customer Satisfaction Quality in Relation to the Assessment of Processes Efficiency - Case Study. Quality – Access to Success. 19(164), 55-59.

Davenport, T., Guha, A., Grewal, D., & Bressgott, T. (2020). How artificial intelligence will change the future of marketing. Journal of the Academy of Marketing Science, 48, 24-42.

Gleim, M. R., Smith, J. S., Andrews, D., & Cronin, J. J. (2013). Against the Green: A Multi-method Examination of the Barriers to Green Consumption. Journal of Retailing, 89(1), 44–61.

Huang, M.-H., & Rust, R. T. (2021). A strategic framework for artificial intelligence in marketing. Journal of the Academy of Marketing Science, 49, 30-50.

Hughner, R. S., McDonagh, P., Prothero, A., Shultz, C. J., & Stanton, J. (2007). Who are organic food consumers? A compilation and review of why people purchase organic food. Journal of Consumer Behaviour, 6(2–3), 94–110.

Chekima**,** B., Bouteraa, M., Ansar, R., Lada, S., Fook, L. M., Tamma, E., Abdul Adis, A.-A., & Chekima, K. (2023). Determinants of Organic Food Consumption in Narrowing the Green Gap. Sustainability, 15(11), 8554.

Lemon, K. N., & Verhoef, P. C. (2016). Understanding customer experience throughout the customer journey. Journal of Marketing, 80(6), 69–96.

Longoni, C., Bonezzi, A., & Morewedge, C. K. (2019). Resistance to medical artificial intelligence. Journal of Consumer Research, 46(4), 629–650.

Magnusson, M. K., Arvola, A., Hursti, U.-K. K., Åberg, L., & Sjödén, P.-O. (2003). Choice of organic foods is related to perceived consequences for human health and to environmentally friendly behaviour. Appetite, 40(2), 109–117.

Paschen, U., Pitt, C., & Kietzmann, J. (2020). Artificial intelligence: Building blocks and an innovation typology. Business Horizons, 63(2), 147–155.

Rana, J., & Paul, J. (2017). Consumer behavior and purchase intention for organic food: A review and research agenda. Journal of Retailing and Consumer Services, 38, 157-165.

Sheth, J. N., Newman, B. I., & Gross, B. L. (1991). Why we buy what we buy: A theory of consumption values. Journal of Business Research, 22(2), 159-170.

Wintschnig, B.A. (2021). The Attitude-Behavior Gap – Drivers and Barriers of Sustainable Consumption, Junior Management Science, 6(2), 324-346.

Yadav, R., & Pathak, G.S. (2016). Young consumers' intention towards buying green products in a developing nation: Extending the theory of planned behavior. Journal of Cleaner Prouction, 135, 732-739.

Zhuo, Z., Ren, Z., & Zhu, Z. (2022). Attitude-Behavior Gap in Green Consumption Behavior: A Review. Journal of Economics, Management and Trade, 28(12), 12-28.

THE IMPACT OF EMAIL MARKETING ON THE PERFORMANCE OF E-COMMERCE BUSINESSES IN THE DIGITAL ECONOMY

Adela Feranecová, Alena Nováková

Bratislava University of Economics and Business, Slovakia
Faculty of Business Economics with seat in Kosice
adela.feranecova@euba.sk, alena.novakova@euba.sk

DOI: 10.35011/IDIMT-2026-359

Keywords

Digital marketing; email marketing; e-commerce; personalization; segmentation.

Abstract

In the world of digital marketing, email marketing is one of the most effective tools for building relationships with both existing and potential customers and for increasing sales. It enables companies to communicate directly with their target audience, personalize the content they send, and measure the effectiveness of email campaigns. Despite rapidly changing trends in online marketing, email marketing maintains a strong position and offers numerous advantages due to its low cost, high return on investment (ROI), and wide possibilities for automation. This paper analyses the impact of email marketing on the performance of e-commerce businesses and explores the factors that determine its effectiveness. The aim of the paper is to examine the influence of email marketing on e-commerce performance and to identify the key factors that determine its effectiveness. The study adopts a conceptual approach based on a review and synthesis of existing theoretical and empirical literature related to digital marketing and consumer behaviour. The findings of the literature review suggest that personalization, segmentation, marketing automation, and data analytics play a crucial role in enhancing customer engagement, conversion outcomes, and overall e-commerce performance.

1. Introduction

The digital economy has fundamentally transformed the way businesses approach customers. Communication has shifted from one-way and often costly forms of advertising to an environment that is fast, measurable, and highly flexible. Within this space, email marketing maintains a particularly strong position, as it combines low costs with a high potential for reaching, engaging, and retaining customers. As indicated in the underlying literature, email marketing is perceived as a tool for building customer relationships, supporting sales, personalizing communication, and monitoring campaign effectiveness.

Its importance in e-commerce lies primarily in its ability to reach customers at the right time with relevant content (Janoskova, 2023). While social media platforms often operate in an environment saturated with content, email is delivered directly to the subscriber’s inbox, creating space for a more

personalized form of communication. This is especially important for e-commerce businesses, which need not only to attract new customers but also to encourage repeat purchases, increase the average order value, and build customer loyalty.

Digital marketing represents a dynamically evolving field that fundamentally transforms communication between businesses and customers (Chaffey & Ellis-Chadwick, 2019; Kotler et al., 2021). As a result of digitalization and technological advancement, marketing activities are increasingly shifting into the online environment (Tiago & Veríssimo, 2014). Email marketing is considered one of the most effective tools of digital communication, particularly due to its capabilities for personalization, automation, and precise performance measurement (Järvinen & Taiminen, 2016; Wedel & Kannan, 2016). In the e-commerce environment, it plays a significant role in managing the customer journey and building long-term customer relationships (Lemon & Verhoef, 2016).

The aim of this paper is to analyse the impact of email marketing on the performance of e-commerce businesses and to identify the key factors that determine its effectiveness. This study is conceptual in nature and is based on a synthesis of existing theoretical and empirical literature on email marketing and e-commerce performance. The paper integrates current knowledge and proposes a conceptual framework explaining the relationship between email marketing practices and e-commerce performance.

In line with this objective, the study addresses the following research question:

- *How does email marketing influence the performance of e-commerce businesses, and which factors determine its effectiveness in the digital economy?*

2. Literature review

2.1. Email marketing as a tool for building customer relationships

Email marketing is a form of digital marketing that enables the delivery of promotional, informational, or transactional messages to a selected segment of contacts. Its main advantage lies in direct communication with customers, which can be precisely tailored to their needs, behavior, and purchase history. The underlying literature emphasizes that email marketing is one of the most effective tools for building customer relationships and increasing sales, with key strengths including low costs, high return on investment, and extensive automation capabilities. Email marketing represents a direct marketing tool that allows firms to communicate with customers through personalized messages (Ryan, 2016). Due to its low cost and high return on investment, it is considered one of the most effective marketing tools.

Modern approaches to email marketing are closely linked to marketing automation and the use of data analytics (Järvinen & Taiminen, 2016). The use of big data enables companies to better understand customer behavior and optimize their marketing activities (Hofacker et al., 2016).

Contemporary literature suggests that the effectiveness of email marketing is not determined by isolated tools, but rather by their integration within a data-driven marketing ecosystem. A key common theme across studies is the emphasis on personalization and data utilization as primary sources of competitive advantage. The synthesis of key theoretical approaches and empirical findings is systematically presented in Table 1.

Table 1. Synthesis of key theoretical approaches and empirical findings

Authors	Main finding	Impact on research/ practice
Lemon & Verhoef (2016)	Customer journey is a multidimensional process	The importance of email marketing in different stages
Kannan & Li (2017)	Digital marketing is data-driven	Support for KPIs and analytics
Bleier & Eisenbeiss (2015)	Personalization increases conversions	The necessity of personalization
Wedel & Kannan (2016)	Analytics improves marketing	Campaign optimization
Järvinen & Taiminen (2016)	Automation increases efficiency	Use of automation
Kumar (2018)	CLV (Customer Lifetime Value) is a key metric	Long-term benefit
Gupta & Lehmann (2003)	Customers are assets	Customer retention
Reinartz & Kumar (2003)	Loyalty increases profit	Email as a retention tool
Ansari & Mela (2003)	Personalization increases effectiveness	Database segmentation
Anderl et al. (2016)	Analysis of the customer journey	Better targeting
Li & Kannan (2014)	Attribution models	Multi-channel marketing
Srinivasan et al. (2016)	Digital channels influence sales	E-commerce growth
Berger & Milkman (2012)	Emotions increase engagement	Email content
Hofacker et al. (2016)	Big data improve decision-making	Data-driven marketing
Dwivedi et al. (2021)	AI is changing marketing	Future development
Payne & Frow (2005)	CRM improves relationships	Email in CRM

Source: own elaboration based on the listed literature

While Lemon and Verhoef (2016) emphasize the contextual and multidimensional nature of the customer journey, Kannan and Li (2017) highlight the growing role of data analytics as a mechanism for optimizing marketing interventions. The integration of these approaches allows email marketing to be understood not only as a linear tool influencing conversions, but as an adaptive system whose effectiveness is conditioned by the stage of the customer journey and the organization's ability to leverage real-time data.

Research by Bleier and Eisenbeiss (2015) and Ansari and Mela (2003) consistently confirms the positive impact of personalization on customers' conversion behavior. However, these approaches implicitly assume the availability of high-quality data and the ability to process it effectively. In this context, data analytics becomes crucial (Wedel & Kannan, 2016; Hofacker et al., 2016), as it enables the transformation of customer data into actionable insights. Without this analytical layer, personalization remains more of a formal than a truly effective tool.

Building on analytical approaches is marketing automation (Järvinen & Taiminen, 2016), which represents a mechanism for operationalizing data-driven decisions. Automation enables the scaling of personalization and real-time responses to customer behavior, thereby linking the strategic and operational levels of marketing. These findings suggest that the true value of email marketing emerges only through the synergistic interaction of analytics, personalization, and automation.

From a long-term perspective, email marketing should also be viewed through the lens of customer value. The concept of Customer Lifetime Value (Kumar, 2018), together with the perspective of customers as assets (Gupta & Lehmann, 2003), shifts performance evaluation from short-term metrics toward long-term profitability. However, Reinartz and Kumar (2003) point out that customer loyalty is not always linearly linked to profitability, suggesting the need for a differentiated approach to customer segment management.

Another important aspect is the complexity of the customer journey. Anderl et al. (2016) and Li and Kannan (2014) highlight that the effectiveness of email marketing cannot be evaluated in isolation, but only within a multi-channel environment and attribution models. This challenges traditional approaches based on simple KPIs (e.g., open rate or CTR) and emphasizes the need for more sophisticated metrics capturing the true contribution of individual channels.

From a content perspective, Berger and Milkman (2012) emphasize the importance of emotional factors that influence customer engagement. However, this aspect is underrepresented in many quantitatively oriented studies, representing a potential research gap.

Finally, technological development (Dwivedi et al., 2021) indicates a shift toward the use of artificial intelligence and predictive analytics, enabling more precise targeting and campaign optimization. At the same time, the importance of CRM systems is increasing (Payne & Frow, 2005), as they integrate various data and communication flows. This trend points to a shift from isolated marketing tools toward integrated customer relationship management platforms.

Overall, the literature agrees on the importance of email marketing as an effective tool, while also revealing its increasing complexity. Effectiveness is no longer determined solely by the use of email as a communication channel, but by organizations' ability to integrate data, technologies, and customer-centric approaches into a unified strategic framework.

Nevertheless, the literature also points to several limitations. The effectiveness of email marketing may decline due to customer fatigue, excessive communication frequency, and increasing competition for consumer attention. Moreover, concerns regarding privacy protection and regulatory requirements such as the General Data Protection Regulation (GDPR) may limit the extent to which organizations can utilize customer data for personalization purposes. These challenges indicate that successful email marketing requires a balance between personalization effectiveness and ethical data management practices.

Despite the extensive literature on email marketing, several research gaps remain. Existing studies frequently focus on isolated aspects such as personalization, automation, or customer retention. However, fewer studies provide an integrated perspective linking email marketing factors, customer engagement, conversion behavior, and business performance within a single conceptual framework. Furthermore, the increasing role of artificial intelligence and predictive analytics in email marketing remains insufficiently explored in the context of e-commerce performance measurement. This paper seeks to address these gaps by proposing an integrated conceptual model based on contemporary literature.

2.2. The impact of Email marketing on performance indicators

Email marketing represents a significant tool within the digital marketing mix, enabling the effective influence of e-commerce performance, particularly through sales growth, customer relationship building, and marketing cost optimization. Its impact is evident across the entire customer lifecycle—from acquisition to retention and repeat purchases.

In the area of customer acquisition, email marketing enables targeted communication with potential customers based on segmentation and behavioral data, thereby increasing the likelihood of conversion. Personalized content, incentives (e.g., discounts), and relevant communication significantly contribute to transforming prospects into actual buyers.

Equally important is its role in customer retention, where regular and personalized communication supports the development of loyalty and long-term relationships. By leveraging customer data,

businesses can deliver relevant offers, thereby increasing the likelihood of repeat purchases and the growth of Customer Lifetime Value.

Industry reports frequently identify email marketing as one of the channels with the highest return on investment, although reported values vary significantly depending on sector, campaign design, and measurement methodology (Donnelly, 2024).

Optimized campaigns also support the growth of AOV through cross-sell and upsell strategies, which use customer data and automation to offer complementary or higher-value products. Similarly, email marketing effectively reduces cart abandonment rates, with automated campaigns achieving high open rates and significantly contributing to purchase completion.

Overall, it can be concluded that email marketing has a significant impact on key e-commerce performance indicators, particularly conversions, revenue, and customer value (Gupta & Lehmann, 2003; Reinartz & Kumar, 2003). Regular communication also supports customer loyalty and increases the rate of repeat purchases (Srinivasan et al., 2016).

3. Results

The proposed conceptual model identifies the fundamental relationships between email marketing factors, customer behavior, and e-shop performance.

The model consists of four main levels. The first level includes email marketing factors, such as segmentation, personalization, email content, and campaign timing. These factors influence the level of customer engagement, which is measured using indicators such as open rate and CTR. Increased customer engagement leads to a positive impact on conversion behavior, which is reflected in a higher conversion rate and completed purchases. In the final stage, these effects are translated into e-shop performance, represented by revenue, return on investment, and customer lifetime value.

The proposed conceptual model systematically explains the mechanism through which email marketing tools influence the performance of e-commerce entities, emphasizing the mediating role of customer behavior.

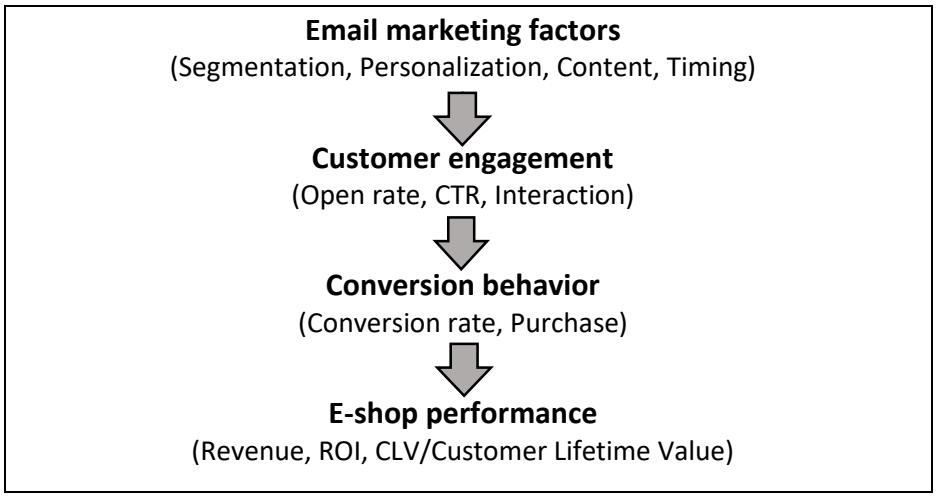


Figure 1. Conceptual model of the impact of Email marketing on e-shop performance Source: authors

The model is based on the assumption that the effectiveness of email marketing is not direct but is mediated by customer engagement and conversion behaviour.

At the first level, four key dimensions of email marketing play a crucial role:

- Segmentation increases communication relevance by reducing information noise.
- Personalization strengthens the perception of individual approach and trust.
- Email content influences both cognitive and emotional customer responses.
- Timing determines the likelihood of opening and response.

These factors collectively determine customer engagement, which represents a critical mediator between marketing activity and economic outcomes. Higher engagement is empirically reflected in the growth of metrics such as open rate and CTR, but especially in the quality of interaction (e.g., time spent with content, clicks through to the e-shop).

Subsequently, engagement influences conversion behaviour, which represents the transition from interest to action. This is a behavioural transformation where a purchase or other desired action takes place. This stage reflects the effectiveness of the marketing funnel.

The final level of the model represents e-shop performance, operationalized through:

- Short-term metrics (revenue, conversion volume).
- Efficiency indicators (ROI).
- Long-term customer value (CLV).

The proposed model assumes that the impact of email marketing on e-shop performance is indirect and occurs through a sequential mechanism of customer engagement and conversion behaviour. This approach reflects the behavioural nature of digital marketing, where economic outcomes are the result of a gradual transformation of customer responses.

4. Discussion

Based on the analysed theoretical approaches and empirical findings, it can be concluded that email marketing represents a stable yet dynamically evolving tool of digital marketing. Its importance in the e-commerce environment is growing, mainly due to the increasing need for personalized communication and effective customer experience management (Lemon & Verhoef, 2016).

A key finding is that the effectiveness of email marketing is conditioned by organizations' ability to work with data and implement advanced analytical tools (Wedel & Kannan, 2016). Personalization and segmentation appear to be key determinants of success, and their importance is supported by several empirical studies (Bleier & Eisenbeiss, 2015; Ansari & Mela, 2003).

The discussion also highlights the need to integrate email marketing into a broader marketing ecosystem. As noted by Anderl et al. (2016), the customer journey is complex and includes multiple touchpoints, with email marketing representing one of the most important communication channels.

From the perspective of future development, an increasing role of artificial intelligence, automation, and predictive analytics can be expected (Dwivedi et al., 2021). These technologies will enable even more precise targeting of marketing campaigns and further improve their effectiveness.

On the other hand, it is also necessary to consider challenges such as customer saturation with marketing messages and increasing requirements for personal data protection (Martin & Murphy, 2017). These factors may affect the future effectiveness of email marketing and require a strategic approach from businesses.

5. Conclusion

Email marketing represents an important tool of digital marketing that has a significant impact on e-shop performance. Based on theoretical and empirical findings, it can be concluded that its effectiveness is primarily conditioned by the use of data, communication personalization, and continuous optimization of marketing activities.

A key benefit of email marketing is its ability to influence the entire customer journey, from acquisition to retention. Properly designed campaigns can increase conversions, revenue, and customer lifetime value, thereby contributing to the long-term competitiveness of businesses.

At the same time, it is evident that the future of email marketing will be closely linked to technological progress, particularly in the areas of artificial intelligence and marketing automation. Businesses that are able to effectively implement these tools will be better positioned to achieve improved results and adapt to the dynamically changing conditions of the digital market.

Acknowledgement

This contribution was developed as part of the project "Analysis of the development of prices of building materials and construction work in Slovakia."

References

Anderl, E., Becker, I., von Wangenheim, F., & Schumann, J. H. (2016). Mapping the customer journey: Lessons learned from graph-based online attribution modeling. International Journal of Research in Marketing, 33(3), 457-474. https://doi.org/10.1016/j.ijresmar.2016.03.001

Ansari, A., & Mela, C. F. (2003). E-customization. Journal of Marketing Research, 40(2), 131-145. https://doi.org/10.1509/jmkr.40.2.131.19224

Berger, J., & Milkman, K. L. (2012). What makes online content viral? Journal of Marketing Research, 49(2), 192-205. https://doi.org/10.1509/jmr.10.0353

Bleier, A., & Eisenbeiss, M. (2015). Personalized online advertising effectiveness. Marketing Science, 34(5), 669-688. https://doi.org/10.1287/mksc.2015.0930

Dwivedi, Y. K., et al. (2021). Artificial intelligence (AI): Multidisciplinary perspectives. International Journal of Information Management, 57, 101994. https://doi.org/10.1016/j.ijinfomgt.2019.08.002

Gupta, S., & Lehmann, D. R. (2003). Customers as assets. Journal of Interactive Marketing, 17(1), 9-24. https://doi.org/10.1002/dir.10045

Hofacker, C. F., Malthouse, E. C., & Sultan, F. (2016). Big data and consumer behavior. Journal of Business Research, 69(2), 897-904. https://doi.org/10.1016/j.jbusres.2015.07.001

Janoskova, M. (2023). E-Commerce: A prosperous form of trading in Slovakia. Acta Oeconomica Cassoviensia, 16(1-2), 25-41.

Järvinen, J., & Taiminen, H. (2016). Harnessing marketing automation for B2B content marketing. Industrial Marketing Management, 54, 164-175. https://doi.org/10.1016/j.indmarman.2015.07.002

Kannan, P. K., & Li, H. A. (2017). Digital marketing: A framework. International Journal of Research in Marketing, 34(1), 22-45. https://doi.org/10.1016/j.ijresmar.2016.11.006

Kotler, P., Kartajaya, H., & Setiawan, I. (2021). Marketing 5.0: Technology for humanity. Wiley.

Kumar, V. (2018). Customer lifetime value. Now Publishers (Foundations and Trends in Marketing). https://doi.org/10.1561/1700000014

Lemon, K. N., & Verhoef, P. C. (2016). Understanding customer experience. Journal of Marketing, 80(6), 69-96. https://doi.org/10.1509/jm.15.0420

Li, H., & Kannan, P. K. (2014). Attribution modeling. Journal of Marketing Research, 51(1), 40-56. https://doi.org/10.1509/jmr.13.0050

Martin, K. D., & Murphy, P. E. (2017). Data privacy in marketing. Journal of the Academy of Marketing Science, 45(2), 135-155. https://doi.org/10.1007/s11747-016-0495-4

Payne, A., & Frow, P. (2005). CRM framework. Journal of Marketing, 69(4), 167-176. https://doi.org/10.1509/jmkg.2005.69.4.167

Reinartz, W. J., & Kumar, V. (2003). Customer relationship characteristics. Journal of Marketing, 67(1), 77-99. https://doi.org/10.1509/jmkg.67.1.77.18589

Srinivasan, S., Rutz, O., & Pauwels, K. (2016). Paths to and off purchase. Journal of the Academy of Marketing Science, 44(4), 440-453. https://doi.org/10.1007/s11747-015-0459-z

Tiago, M. T. P. M. B., & Veríssimo, J. M. C. (2014). Digital marketing and social media. Business Horizons, 57(6), 703-708. https://doi.org/10.1016/j.bushor.2014.07.002

Wedel, M., & Kannan, P. K. (2016). Marketing analytics. Journal of Marketing, 80(6), 97-121. https://doi.org/10.1509/jm.15.0413

CONSUMER BEHAVIOUR OF GEN Z IN THE AGE OF AI

Jaromír Veber, Petra Králová

VSCHT, School of Business, Czech Republic
jaromir.veber@vscht.cz, petra.kralova@vscht.cz

DOI: 10.35011/IDIMT-2026-367

Keywords

AI; consumer behaviour; social networks; personalisation; sustainability; streaming platforms.

Abstract

This article examines the consumer behaviour of Generation Z in the context of digital marketing, purchasing incentives, and marketing tools used in the age of artificial intelligence. It analyses how selected digital purchasing incentives, especially personalised offers, social media advertising, FOMO messages, free shipping, and subscription services, influence young consumers' willingness to purchase and how these incentives relate to sustainable consumer behaviour. The empirical section is based on a questionnaire survey of 220 Generation Z respondents. The results show that respondents are not merely passive recipients of digital marketing stimuli: when purchasing food and clothing, they most often emphasise personal experience, price, quality, and need. At the same time, some incentives encourage consumption, especially free shipping and, to a lesser extent, subscription services. Because respondents usually cannot distinguish AI personalisation from broader digital marketing, the article interprets the findings as preliminary research into online consumer behaviour in the age of AI.

1. Current trends

Marketing plays an important role in informing current and potential customers about products and related developments. Its role has gradually shifted from basic information, through the strengthening of competitive advantage, to more sophisticated and sometimes ethically questionable forms of influencing consumer behaviour. (Kotler & Keller, 2013; Kotler et al., 2025) Digitalisation and artificial intelligence have further intensified this shift by allowing promotional campaigns to be targeted at precisely defined consumer groups and, in some cases, individual users.

In this article, we distinguish between three levels of marketing impact. Digital marketing refers to a broader set of online activities, including social media advertising, email marketing, content marketing, and price incentives in e-shops. Digital shopping incentives refer to specific stimuli that accelerate or reinforce purchasing decisions, such as free shipping, time-limited offers, product recommendations, and discounted subscriptions. AI marketing is a narrower category that uses algorithmic personalisation, recommendation systems, predictive models, dynamic pricing, or personalised content. Since consumers often cannot identify whether a message is produced by AI personalisation, paid advertising, or standard online content, the empirical section focuses on digitally mediated shopping incentives in the age of AI.

AI-powered algorithms identify what customers are interested in, what they buy, and how they shop. Such systems may create conditions for more frequent or more impulsive purchases, although the strength of this effect depends on the type of incentive, product category, and consumer context. The key issue is whether algorithms primarily recommend products that customers need or whether they stimulate additional demand through personalisation, discounts, dynamic pricing, trends, and short product lifecycles.

The relationship between AI-based marketing and sustainable consumption is therefore ambivalent. On the one hand, AI can simplify and accelerate consumption, thereby increasing the volume of products, packaging, transport, and waste. On the other hand, AI can also help customers make more responsible decisions by recommending more durable, repairable, local, or environmentally friendly alternatives, optimising energy consumption, or helping reduce food waste.

The article aims to analyse how selected digital shopping incentives influence the consumer behaviour of Generation Z and whether they contribute to increased consumption or to more responsible purchasing decisions. It addresses three research questions: what factors respondents identify as most important when purchasing groceries, clothing, and subscription services; how they react to selected digital shopping incentives; and what role sustainability plays in their purchasing decisions.

2. The pros and cons of marketing

Advertising and marketing have a long-standing tradition linked to the development of modern business and industrialisation. As production expanded and supply began to exceed demand, companies had to compete not only by price but also by quality, delivery flexibility, responsiveness to customer needs, and differentiation. (Porter, 2008) The development of marketing strategies, benchmarking, and digital communication gradually expanded the means by which firms influence customers.

With the rise of digitalisation, marketing increasingly moved from traditional media to websites, e-shops, platforms, and social networks. This transition brought new possibilities for targeting and measurement, but also strengthened practices such as brainwashing, greenwashing (Delmas & Burbano, 2011), and the stimulation of unnecessary consumption. In the context of AI, this development raises new questions about how far marketing can go in shaping individual choices and whether consumers, especially members of Generations Z and Alpha (Dolot, 2018; Schewe & Meredith, 2004), can maintain distance from personalised messages.

Modern marketing thus has both positive and negative effects. It supports sales, production, and organisational performance, but it can also contribute to consumerism and the normalisation of superfluous purchasing. The following section, therefore focuses on consumer society and its relation to health, lifestyle, and sustainability.

3. Consumer society, health and sustainability

The term consumer society refers to a social environment in which consumption becomes a key value and a measure of individual success. This article does not deny the economic function of marketing, but it emphasises that some marketing activities may also lead to unnecessary or undesirable effects. Two areas are particularly relevant: the impact on health and lifestyle, and the impact on the environment.

Food consumption illustrates this tension clearly. In most developed societies, food availability is no longer the main problem; rather, consumers often face an excess of options, including highly processed and convenience foods. This can contribute to food waste, overeating, and a preference for products that are difficult to associate with a healthy lifestyle. According to the National Institute of Public Health (SZÚ, 2024), the Czech population reports high levels of overweight and obesity, including among children. Although this article does not directly examine health outcomes, food purchasing is included because digital marketing and convenience incentives may influence everyday choices.

The environmental consequences of consumer society are equally important. The Club of Rome report The Limits to Growth (Meadows et al., 1972) and later the concept of sustainable development (United Nations, 2002; World Commission on Environment and Development, 1987) highlighted the tension between economic growth and environmental protection. Overproduction, hoarding, the discarding of unsold fashion items, packaging, transport, and waste all increase pressure on natural resources and ecosystems.

In the digital marketing landscape, these environmental impacts can be intensified by specific purchasing incentives. Free shipping can encourage more frequent or larger orders; time-limited messages can stimulate impulse purchases; fast fashion shortens product lifecycles; and subscription services can maintain long-term digital consumption that is less visible than physical consumption. At the same time, digital and AI tools can have the opposite effect if they recommend more durable, repairable, local, or environmentally friendly options. This dual potential makes it important to examine not only general consumption but also the mechanisms that encourage more frequent or less conscious purchases.

4. Methodology

The research was conducted as a questionnaire survey of 220 Generation Z respondents, primarily high school and college students. Because the sample focused on this specific group, the results cannot be generalised to the entire Generation Z. The survey captured attitudes and reported behaviour regarding food, clothing, subscription services, and more sustainable product options.

The questionnaire included demographic characteristics, consumption habits, the influence of the social and online environment, reactions to selected purchasing incentives, and attitudes toward sustainability. Digital purchasing incentives were examined through reactions to FOMO messages, free shipping, advertising, and subscription offers. Sustainable consumer behaviour was examined through preferences for more sustainable product options, interest in secondhand clothing, preference for higher-quality materials, and the importance of functionality and value for money.

The data were evaluated using descriptive analysis of frequencies and percentages. Selected responses were interpreted in relation to gender, nationality, and type of residence. The study is exploratory. A limitation is the specific composition of the sample: most respondents live in cities with more than one million inhabitants, which may influence service availability and exposure to digital marketing. Another limitation is the cultural heterogeneity of the sample. This is also a strength because it captures a broader range of consumer experiences; however, subgroup sizes do not allow reliable generalisations by nationality. Finally, the research did not directly distinguish whether a specific purchase trigger was caused by AI personalisation, paid advertising, or standard online content. The conclusions therefore apply to digital purchase incentives in the age of AI, not exclusively to demonstrably AI-driven tools.

5. Findings and results

The findings are organised into four categories: food and clothing purchases, subscription services, the influence of social and online environments, and sustainability and responses to purchasing incentives. The interpretation focuses on whether respondents appear to make rational decisions or are sensitive to digital marketing stimuli.

The sample consisted of 85 men (nearly 39%), 130 women (59%), and 5 respondents (2%) who declined to disclose their gender. The majority of respondents were aged 18–20 (121 respondents, 55%), while 67 respondents (30.5%) were aged 21–25. Most respondents live in a city with more than one million inhabitants (172 respondents, 78.2%), and 160 respondents (72.7%) reported completing high school with a diploma, typically as current university students. The sample was culturally heterogeneous: 70 respondents (31.8%) identified as Czech, 42 (19%) as Ukrainian, and 37 (16.8%) as Turkish; other nationalities included Kazakh, Ecuadorian, Russian, Indian, and Chinese. Nearly one third had lived in the Czech Republic for less than one year. Almost half of respondents (44%) had some work experience and income; 40% lived in a dormitory and 20% with their parents.

5.1. Selection of food and clothing

The research examined the influence of the online environment, especially social media and advertising. Although these influences may be algorithmically driven, the study did not distinguish whether respondents were directly influenced by AI personalisation, paid advertising, or regular social media content. The results show that 60% of respondents perceive their own experience and personal opinion as the main decision-making factor; 22% cite recommendations from colleagues and acquaintances; and 15% state that social media content influences their decisions. If the 3% who mention advertising are added, marketing communication accounts for 18% of responses overall.

This suggests that respondents do not perceive advertising and social media as having a dominant direct influence. However, indirect digital influence cannot be ruled out because personal experience, brand preferences, and perceptions of quality may be shaped over time by online content, recommendation algorithms, and repeated exposure to marketing messages. The results therefore point to a combination of rational decision-making and indirect digital influence rather than the clear dominance of marketing communication.

When choosing food, respondents primarily consider price and ingredients, including additives, salt and sugar content, vitamins, and nutritional value. Seventy per cent prefer fresh vegetables and fruit and can prepare simple meals themselves, while 30% prefer semi-prepared foods, ready-made meals, or fast food. Overall, food purchasing appears to be relatively independent: in 83% of cases, respondents actively purchase groceries for themselves or their loved ones, while in 17% of cases groceries are purchased by parents or a partner.

Clothing purchases show a similar pattern. Eighty-two per cent of respondents indicated that they primarily consider their own needs, personal experience, financial situation, and opinions of close colleagues and acquaintances. Some respondents emphasised secondhand clothing, sustainable fashion, high-quality materials, and longevity. Eighteen per cent mentioned the influence of advertising and social media. Overall, clothing purchases were predominantly practical and personally driven rather than primarily trend-driven or marketing-driven.

5.2. Subscription services

Subscriptions can drive consumption by binding customers into long-term relationships with platforms and service providers. The research therefore focused on the rationality of subscription use. Fifty-four per cent of respondents stated that they have multiple subscription services and regularly assess their usage, cancelling unused ones. Twenty-seven per cent reported having more subscription services than they actively use, 13% were considering purchasing a subscription, and 6% did not have any such service.

Among respondents who have subscriptions, 56% pay for 1–3 subscriptions, 21% for 4–5 subscriptions, and 10% for more than five subscriptions. The most popular subscriptions include streaming platforms for movies and music (65% of responses), cloud storage (22%), artificial intelligence services such as ChatGPT Plus and Gemini (8%), and Duolingo (5%). Gaming and creative apps appeared only rarely.

Service-based subscriptions such as Alza Plus or Rohlík Extra were less attractive. Twenty per cent of respondents have such a subscription and actively use it, 5% have one but do not use it, 9% are interested in this type of subscription, and 66% are not considering it. These results suggest that many respondents manage their subscriptions rationally, but the share of unused or underused subscriptions indicates the risk of inertial consumption in digital environments.

5.3. The influence of the social and online environment

Across purchases, 50% of respondents make decisions independently of their community's opinions or are only marginally influenced by them, especially regarding fashion accessories. Thirty-five per cent rely on their parents' opinions, while 15% care about how their community perceives the products they purchase. Family influence thus remains relevant, possibly because of respondents' age, financial support, or established shopping habits.

The survey also examined whether respondents share brands or products on social media that reflect their values. Only 9% said that their online presence is part of their identity and that they enjoy doing so (Jensen Schau & Gilly, 2003). Another 10% reported doing so only when they are in a bad mood, 19% said that they do not present themselves on social media at all, and 36% separate the presentation of consumption from the presentation of themselves as individuals. The culturally heterogeneous sample suggests that consumer behaviour cannot be explained solely by generation; cultural experience, length of residence in the Czech environment, and social background may also play a role (De Mooij & Hofstede, 2011).

5.4. Sustainability and purchasing incentives

Questions about sustainable product offerings showed that 32% of respondents feel moral satisfaction when purchasing a more sustainable version of a product, while 8% are inspired to buy such a version if friends, family, or influencers do so. However, 60% primarily evaluate functionality and value for money and pay less attention to sustainable aspects. Sustainability is therefore not the main purchasing criterion for most respondents, and price can be an important barrier to more sustainable consumption.

Responses to the FOMO message “while supplies last” (Przybylski et al., 2013) suggest a relatively sceptical approach. Twenty-six per cent of respondents would hurry to purchase, 39% perceive the message as a sales gimmick, and 35% shop independently of this incentive. Only a limited number of respondents therefore feel strong pressure to buy quickly.

Free shipping has a stronger effect. Twenty-two per cent of respondents do not worry about shipping costs and are willing to pay for them. However, 46% consider free shipping a strong motivator to buy, and within this group 8% choose purchases based on whether shipping is free and will not buy without it. Another 32% state that free shipping motivates them to purchase more items. Compared with FOMO messages, respondents react more strongly to incentives that offer a direct financial benefit. They are not immune to marketing incentives, but their responses are selective and depend on price sensitivity and practical utility.

The study also examined attitudes toward advertising. Most respondents accept advertisements if they receive free content in return; 73% stated this view. Twenty per cent pay for ad-free online content, 2% see this as a way to feel exclusivity, and 7% are unaware of the option to pay for ad-free content.

6. Conclusions

This study shows that digital shopping incentives and marketing tools used in the age of AI influence the consumer behaviour of Generation Z, but this influence is not clearly dominant. When purchasing groceries and clothing, respondents most often emphasise their own experience, personal needs, price, quality, and financial means. The surveyed segment of Generation Z is therefore not merely a passive recipient of personalised advertising, but rather tends to make practical and relatively rational decisions while maintaining some distance from direct marketing communication.

At the same time, some incentives do boost consumption. Free shipping has the greatest impact: 46% of respondents view it as a strong motivator to buy, and 32% report that it increases the number of items purchased. FOMO-based messages have a more limited effect, with 26% saying they would hurry to buy, while a large share either perceive the message as a sales gimmick or shop independently of it. Subscriptions are similarly ambivalent: more than half of respondents monitor and cancel unused subscriptions, but 27% have more subscriptions than they actively use.

Regarding sustainability, the results are mixed. Some respondents view more sustainable product options positively, but for the majority, functionality and value for money remain decisive. Marketing incentives in the age of AI can therefore either support responsible consumption by recommending durable or environmentally friendly options, or boost consumption when used mainly for personalised advertising and sales growth.

The main contribution of this article is the finding that a simple image of Generation Z as passive recipients of personalised advertising is insufficient. Respondents report that experience, price, quality, and functionality are highly important to them, yet specific digital incentives, especially free shipping and subscriptions, influence their decisions. The results should be understood as preliminary research on digitally mediated consumer behaviour in the age of AI, not as direct evidence of the effects of specific AI algorithms. Further research should expand the sample beyond metropolitan areas, distinguish more precisely between standard digital advertising and AI personalisation, and test the identified relationships statistically.

References

De Mooij, M., & Hofstede, G. (2011). Cross-cultural consumer behavior: A review of research findings. Journal of International Consumer Marketing, 23(3–4), 181–192.

Delmas, M. A., & Burbano, V. C. (2011). The drivers of greenwashing. California Management Review, 54(1), 64–87.

Dolot, A. (2018). The characteristics of Generation Z. E-mentor, 74(2), 44–50.

Jensen Schau, H., & Gilly, M. C. (2003). We are what we post? Self-presentation in personal web space. Journal of Consumer Research, 30(3), 385–404.

Kotler, P., & Keller, K. L. (2013). Marketing management (14th ed.). Grada.

Kotler, P., Kartajaya, H., & Setiawan, I. (2025). Marketing 6.0: Budoucnost ve znamení zážitků. Grada.

Meadows, D. H., Meadows, D. L., Randers, J., & Behrens, W. (1972). The limits to growth. Universe Books.

Porter, M. E. (2008). Competitive advantage: Creating and sustaining superior performance. Simon & Schuster.

Przybylski, A. K., Murayama, K., DeHaan, C. R., & Gladwell, V. (2013). Motivational, emotional, and behavioral correlates of fear of missing out. Computers in Human Behavior, 29(4), 1841–1848.

Schewe, C. D., & Meredith, G. (2004). Segmenting global markets by generational cohorts: Determining motivations by age. Journal of Consumer Behaviour, 4(1), 51–63.

Státní zdravotní ústav. (2024). Nadváha a obezita. https://szu.gov.cz/temata-zdravi-a-bezpecnosti/podpora-zdravi/nadvaha-a-obezita/

United Nations. (2002). World Summit on Sustainable Development, Johannesburg, South Africa.

Wilding, J. P., Batterham, R. L., Calanna, S., Davies, M., Van Gaal, L. F., Lingvay, I., … Kushner, R. F. (2021). Once-weekly semaglutide in adults with overweight or obesity. New England Journal of Medicine, 384(11), 989–1002.

World Commission on Environment and Development. (1987). Our common future. United Nations. https://sustainabledevelopment.un.org/content/documents/5987our-common-future.pdf

COMPETENCE DEMAND IN AI-DRIVEN LABOUR MARKET IN SLOVAKIA

Cecília Olexová, Patrik Koba

Bratislava University of Economics and Business
Faculty of Business Economics with seat in Košice, Slovakia
cecilia.olexova@euba.sk, patrik.koba@euba.sk

DOI: 10.35011/IDIMT-2026-375

Keywords

Competences; skills; labour market; online advertisement; Slovakia.

Abstract

The paper analyses the evolution of competence requirements in the Slovak labour market during the period 2019–2024, driven by the diffusion of AI. Utilizing Cedefop data, the research identifies a structural shift from static knowledge to dynamic skills and transversal competencies. Results show that while all categories grew, transversal skills exhibited the highest relative growth with willingness to learn tripling its frequency. The findings underscore a transition towards hybrid competence profiles where learning agility acts as a critical meta-competence for human-machine collaboration. This transformation signals a move to seek individuals with high developmental potential, forming a necessary foundation for building a resilient and sustainable workforce in a dynamic environment under digitalisation.

1. Introduction

The ongoing diffusion of artificial intelligence (AI) and digital technologies represents a significant driver of change in contemporary labour markets. These developments are associated with transformations in work organisation, job design, and, most importantly, competence requirements. Rather than affecting employment only in quantitative terms, AI increasingly reshapes the qualitative structure of labour demand.

At the theoretical level, these changes are reflected in a gradual shift from traditional task-based approaches to job analysis towards more flexible and integrative framework of work design and competence modelling. Traditional approaches to job analysis focus on specific tasks and responsibilities performed by job incumbents; however, they have been criticised for their limited ability to capture the complexity and variability of contemporary work context (Morgeson & Humphrey, 2006; Dierdorff & Morgeson, 2007; Rogiers & Collings, 2024). In particular, research highlights that work roles and their requirements are shaped by broader occupational and organisational context, which challenges static descriptions of jobs. At the same time, existing approaches to competence modelling do not fully reflect the dynamic and rapidly changing nature of skill requirements in AI-driven labour markets. This creates a gap between established theoretical frameworks and the actual evolution of labour demand observed in practice.

In response to these limitations, more comprehensive frameworks, such as the Occupational Information Network (O*NET) or National Classification of Occupations in Slovakia, have been developed to capture a wide range of worker characteristics, including knowledge, skills, abilities and work activities (Peterson et al., 2001). These approaches move beyond narrow tasks descriptions and provide a more dynamic representation of work, which is particularly relevant in the context of technological change (see e.g. Rogiers & Collings, 2024).

The aim of this paper is to analyse the development of competence requirements in the context of artificial intelligence, with a focus on ESCO skill categories and their relevance for sustainable labour market development. Using Cedefop data on online job advertisements for the period 2019-2024 in Slovakia, the paper examines changes in knowledge, skills, and transversal competences and identifies key trends in their development.

2. Theoretical background

The widespread diffusion of AI is driving a systemic transformation towards intelligent and automated processes across industries, enhancing efficiency, reducing human error, and simultaneously reshaping labour market structures and competence requirements (Zhou et al., 2025).

Recent research highlights a shift from static, knowledge-based qualifications towards dynamic, skill-oriented and adaptive competence profiles (Autor et al., 2003; Autor, 2015; Deming, 2017). In this context, the concept of "skill-biased technological change" explains how technological advancement increases demand for higher-order cognitive, social, and digital skills, while routine tasks become mostly automated. AI primarily affects the task structure jobs by expanding the set of tasks that can be performed by algorithms, while simultaneously altering the demand for human labour across these tasks (Acemoglu et al., 2022). From a human resource management (HRM) perspective, this transition reflects a paradigmatic change from job analysis and traditional qualification models towards competency-based frameworks that emphasise flexibility, learning agility, and employability (Campion et al., 2011). Competences are no longer viewed as fixed assets but as evolving capabilities that must be continuously developed in response to technological and organisational change. This is consistent with the dynamic capabilities' theory, which posits that organisations and their employees must constantly reconfigure their competencies to remain competitive in turbulent environments (Teece, 2018). In fast-changing knowledge economy, labour markets prioritise hybrid competence profiles, integrating domain-specific knowledge with transversal skills such as problem-solving, collaboration, and adaptability (Frey & Osborne, 2017; van Laar et al., 2017).

The analysis of competence requirements in AI-driven labour markets can be grounded in several established frameworks. The European Skills, Competences, Qualifications and Occupations (ESCO) framework classification provides a structured taxonomy of knowledge, skills, and competences, enabling systematic analysis of labour market demand across occupations. The concept of digital competence has been operationalised in frameworks such as DigComp, which defines key areas including information processing, communication, problem-solving, and digital content creation (Vuorikari et al., 2022). Furthermore, ESCO is complemented by O*NET-based approaches, widely used in empirical labour market research to analyse task and skill structures (e.g. Frey & Osborne, 2017). The concept of transversal competences is reflected in models emphasising soft skills, social skills, and adaptability, which have been shown to increase in importance in labour markets (Deming, 2017; Poláková et al., 2023).

Evidence from large-scale analyses of online job advertisements further indicates that AI adoption is associated with increased "skill churn", characterised by both the introduction of new skills and the retirement of previously demanded ones, reflecting a dynamic restructuring of competence

requirements (Acemoglu et al., 2022). A particularly significant theoretical construct in this context is the notion of lifelong learning and learning agility. The ability to continuously acquire new skills is a critical meta-competence, often outweighing existing technical expertise, as it can contribute to the development of human resources and competitive advantage of a company (Noe et al., 2014). This is reflected in the rising demand for "willingness to learn" as a transversal competence, signalling a shift in recruitment logic, from selecting fully qualified candidates to identifying individuals with high developmental potential. Such trends can be interpreted as structural changes in the distribution of skill demand, where variance increasingly concentrates around a limited set of high-value competencies. Moreover, the polarisation of competence demand represents another key theoretical insight. While demand for advanced digital and analytical skills grows rapidly, low-skill routine tasks decline, leading to a labour market polarisation (Goos et al., 2014). AI does not merely substitute existing skills but expands the overall skill portfolio required in occupations, enabling firms to define more precise and forward-looking competence requirements (Chen et al., 2026).

The theoretical literature converges on three central arguments:

1. A shift from knowledge to skills and transversal competences,
2. The emergence of hybrid and adaptive competence profiles, and
3. The increasing importance of continuous learning as a core employability factor.

3. Data and methods

The primary data source is the Cedefop Skills-OVATE database, which aggregates real-time labour market information through the analysis of online job advertisements (OJAs). The dataset focuses on Slovakia and provides insights based on the ESCO classification. The analysis covers the period from 2019 to 2024. To provide a broader context, the 2024 data are compared with the European Union average.

To evaluate the dynamics of skill requirements over the period 2019-2024, the aggregate analysis was applied across three primary pillars: Knowledge, Skills, and Transversal Competencies. The Compound Annual Growth Rate (1) was employed as a smoothed annual growth rate that accounts for cumulative changes over time. This metric was used to enable a standardized comparison between different skill categories (Knowledge, Skills, and Transversal competencies), regardless of their baseline values in 2019.

$$CAGR = \left(\frac{V_{t_n}}{V_{t_o}}\right)^{\frac{1}{t_n - t_0}} - 1 \qquad (1)$$

Where:

V_{t_0} represents the initial value in the base year,
V_{t_n} represents the initial value in the target year,
n denotes the number of compounding periods.

Additionally, absolute changes were expressed in percentage points (p.p.) to represent the direct difference between the relative shares of requirements between 2019 and 2024. The 2024 competence profiles for Slovakia were compared against the EU-27 average to identify national specializations or gaps. Subsequently, each competence category was analysed in detail to provide a granular view of the observed trends, and the standard deviation (SD) and coefficient of variation (CV) were employed to measure the dispersion and consistency of requirements.

4. Results

Figure 1. illustrates the development of ESCO skill categories over the period 2019-2024, while Table 1. presents the values for 2019 and 2024 together with the corresponding changes.

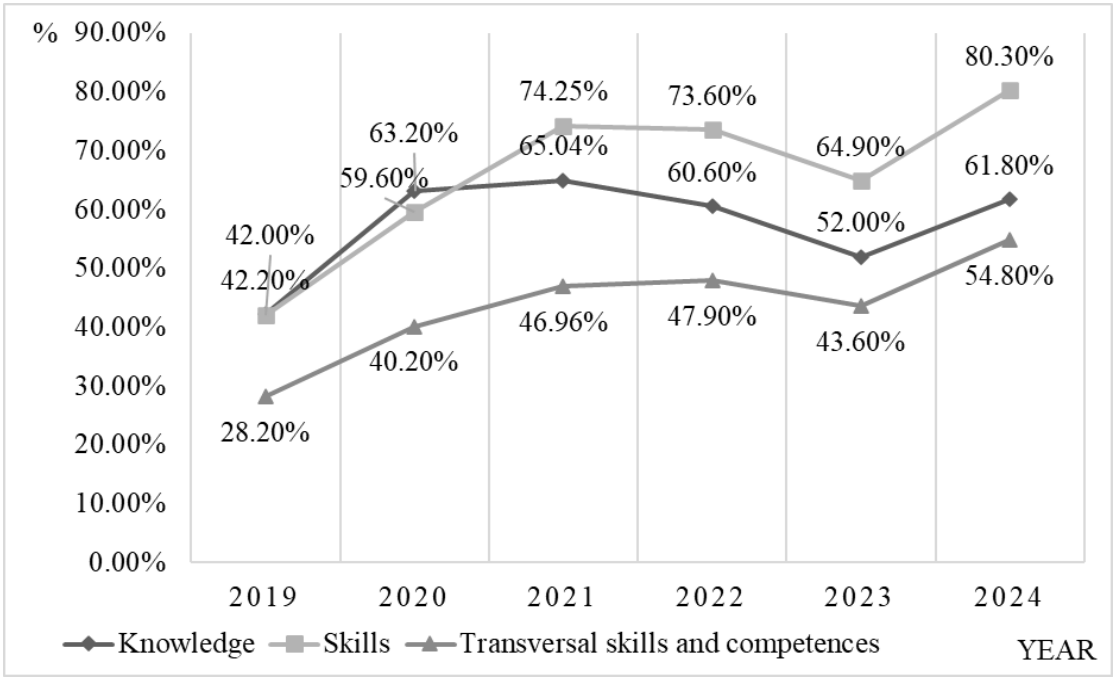


Figure 1. Skills in OJAs (2019-2024) Source: authors´ own processing (data: Cedefop 2019-2024)

Table 1. Quantitative Dynamics of ESCO Skill Requirements (2019, 2024)

Esco skills	**2019 (t_0)**	**2024 (t_n)**	**Abs. Change (Δ)**	**CAGR**
Knowledge	42.20%	61.80%	+19.60 p.p.	7.93%
Skills	42.00%	80.30%	+38.30 p.p.	13.84%
Transversal skills and competences	28.20%	54.80%	+26.60 p.p.	14.21%

Source: authors´ own processing (data: Cedefop 2019, 2024)

The data reveals a shift in competence requirements between 2019 and 2024. The most prominent growth is observed in the Skills category, which saw an increase of 38.30 p.p., reflecting a CAGR of 13.84%. Transversal skills and competences had the highest relative growth rate at 14.21% CAGR, nearly doubling their presence in requirement profiles over the analysed period. In contrast, while knowledge requirements also increased, their growth was notably more moderate. The results indicate a shift from knowledge-based competencies towards skills-based and transversal competencies, highlighting the increasing importance of practical and human-centric abilities in the age of AI. A temporary decline observed in 2023 can be interpreted as a stabilisation effect following the rapid post-pandemic expansion of skill demand, combined with macroeconomic uncertainty affecting labour market dynamics.

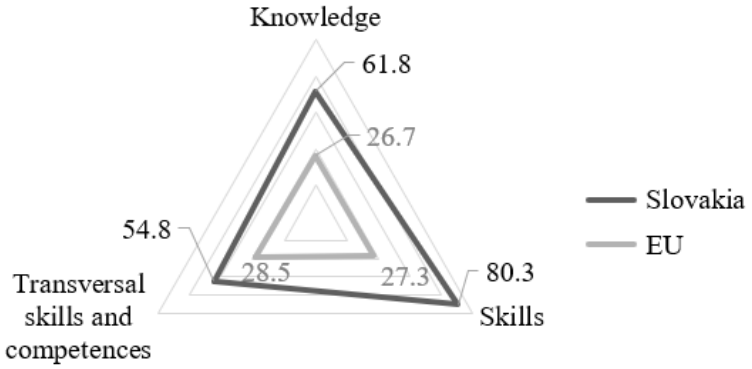


Figure 2. Comparison of skill requirements intensity in 2024 (%) Source: own processing (data: Cedefop 2024)

While the primary objective of this study was to compare Slovakia and the EU-27, the analysis revealed significant structural disparities that warrant further attention (Figure 2.). Specifically, the frequency of competence requirements in Slovakia appears disproportionately higher than the EU-27 average across all monitored categories. These findings suggest that Slovak recruitment documentation is characterised by high granularity and over-specification, whereas the aggregate EU-27 data reflects a more streamlined and purposeful approach to definition of skill requirements.

4.1. Knowledge requirements (2019-2024)

The analysis revealed a broad spectrum of required knowledge domains, which are presented here in descending order of their frequency in 2024: (1) Business and administration, (2) Personal skills and development, (3) ICTs, (4) Engineering and engineering trades, (5) Architecture and construction, (6) Mathematics and statistics, (7) Manufacturing and processing, (8) Health, (9) Languages, (10) Social and behavioural sciences. To maintain focus on the most impactful structural shifts, Table 2 presents the top three knowledge categories.

Table 2. Leading knowledge categories in OJAs (2019-2024)

Knowledge	**2019 (%)**	**2020 (%)**	**2021 (%)**	**2022 (%)**	**2023 (%)**	**2024 (%)**	**SD**	**CV (%)**
Business and administration	20.29	31.1	31.79	26.6	21.4	25.58	4.77	18.26
Personal skills and development	15.69	21.44	23.25	24.04	18.35	22.16	3.19	15.32
ICTs	16.88	23.73	24.26	23.41	18.2	22.07	3.12	14.58

Source: authors´ own processing (data: Cedefop 2019-2024)

As shown in the table, three leading categories are very close in 2024, reflecting the structural demands of the industry. ICT proficiency has transitioned from a specialized skill to broad requirements across all sectors, complemented by a rising demand for personal skills and development.

4.2. Skills requirements (2019-2024)

Required skills are presented here in descending order according to their frequency in 2024 OJAs: (1) Providing information and support to the public and clients, (2) Accessing and analysing digital data, (3) Using digital tools for collaboration, content creation and problem solving, (4) Solving problems, (5) Supervising people, (6) Programming computer systems, (7) Promoting, selling and purchasing, (8) Liaising and networking, (9) Creating artistic, visual or instructive materials, (10)

Protecting and enforcing, (11) Organising, planning and scheduling work and activities, (12) Performing administrative activities, (13) Designing systems and products, (14) Developing objectives and strategies, (15) Communication, collaboration and creativity, (16) Analysing and evaluating information and data, (17) Managing information, and (18) Setting up and protecting computer systems. The top three skill categories required by labour market are presented in Table 3.

Table 3. Leading skills categories in OJAs (2019-2024)

Skills	2019 (%)	2020 (%)	2021 (%)	2022 (%)	2023 (%)	2024 (%)	SD	CV (%)
Providing information and support to the public and clients	3.61	6.45	35.23	35.15	38.43	56.2	20.29	69.54
Accessing and analysing digital data	24.04	34.42	34.74	34.77	26.64	33.47	4.75	15.15
Using digital tools for collaboration, content creation and problem solving	19.49	27.76	29.87	30.63	23.31	28.3	4.30	16.20

Source: authors´ own processing (data: Cedefop 2019-2024)

The data points to a surge in the requirement for Providing information and support to the public and clients, which soared from a marginal 3.61% in 2019 to 56.20% in 2024; this is reflected in the extreme variability (SD = 20.29; CV = 69.54%). This suggests that even as work becomes more digitalized, the ability to manage client relationships and provide professional support has become a highly prioritized standard across almost all sectors. Simultaneously, the high values for Accessing and analysing digital data (33.47%) and Using digital tools (28.30%) reinforce the previously mentioned "universalist" profile.

4.3. Transversal skills and competences requirements (2019-2024)

Transversal requirements are sequenced according to their relative importance in 2024 OJAs: (1) Demonstrating willingness to learn, (2) Taking a proactive approach, (3) Collaborating in teams and networks, (4) Working efficiently, (5) Thinking creatively and innovatively, (6) Working with digital devices and applications, and (7) Leading others. Table 4 presents the top four transversal categories.

Table 4. Transversal skills and competences in OJAs (2019-2024)

Transversal competences	2019 (%)	2020 (%)	2021 (%)	2022 (%)	2023 (%)	2024 (%)	SD	CV (%)
Demonstrating willingness to learn	10.24	14.6	22.36	21.63	23.23	31.23	7.31	35.59
Taking a proactive approach	17.85	25.74	27.44	27.89	21.5	25.9	3.92	16.07
Collaborating in teams and networks	7.02	10.87	12.4	12.25	9.54	11.84	2.07	19.46
Working efficiently	9.03	15.09	12.59	11.92	9.41	11.75	2.22	19.10

Source: authors´ own processing (data: Cedefop 2019-2024)

The most significant finding in this category is the rapid rise of demonstrating willingness to learn, which tripled its frequency from 10.24% in 2019 to 31.23% in 2024. The high variability of this trend (SD = 7.31; CV = 35.59%) confirms that the demand for learning agility is the most dynamic factor among transversal competences.

5. Conclusion

The results indicate a shift in competence demand towards skills and transversal competences within AI-driven labour market. Consistent with the framework of skill-biased technological change, this suggests an increasing importance of adaptability and continuous development reflecting the theory of dynamic capabilities (Teece, 2018). As work is increasingly situated in socio-technical systems, digital interaction, data analysis, and communication-oriented skills make human-machine collaboration the new norm. This transformation aligns with the theory of labour market polarisation (Goos et al., 2014), where a small group of high-value skills dominates the market. Modern employees are expected to be "digital natives" capable of data-driven decision-making, while also possessing hybrid competence profile (van Laar et al., 2017) for effective interaction. As emphasized by Kohnová et al. (2020), the long-term sustainability of the labour market is closely linked to effective job design and the strategic alignment of employee skills with the challenges of Industry 4.0. Consequently, companies are moving away from seeking experts with static knowledge. Instead, there is an increasing demand for individuals with a growth mindset, capable of adapting to new tools and technologies, such as Generative AI. As also noted by Noe et al. (2014), demonstrating willingness to learn acts as a fundamental competence; in the context of our findings, this trait determines the long-term utility and evolution of all other professional skills, contributing to a more resilient and sustainable talent ecosystem within an AI-augmented environment.

These shifting dynamics carry significant practical implications. For employers and HR management, recruitment strategies must pivot toward evaluating candidates' developmental potential and cognitive flexibility rather than static technical qualifications. Consequently, education systems and corporate training programmes need to shift focus from narrow specialization toward cultivating lifelong learning habits to effectively mitigate emerging skills mismatches. This is particularly important in the Slovak context, where strategic employee training within companies are increasingly recognized as a key component for achieving organizational adaptation and long-term competitiveness in a changing economic environment (Smerek & Vetráková, 2020).

Acknowledgement

This research was supported by the project *"Job Redesign and the Revision of the Competency Model in the Era of Artificial Intelligence"* (Contract for the Provision of Financial Resources for Project Financing) and by the project KEGA 020EU-4/2025 *"The Development of a Model for an Innovative Didactic Concept to Enhance Green Skills of Students at Economics-focused Universities with an Emphasis on Ecosystem Sustainability".*

References

Acemoglu, D., Autor, D., Hazell, J., & Restrepo, P. (2022). Artificial intelligence and jobs: Evidence from online vacancies. Journal of Labor Economics, 40(S1), 293-340. https://doi.org/10.1086/718327

Autor, D. H. (2015). Why are there still so many jobs? The history and future of workplace automation. Journal of Economic Perspectives, 29(3), 3-30. https://doi.org/10.1257/jep.29.3.3

Autor, D. H., Levy, F., & Murnane, R. J. (2003). The skill content of recent technological change: an empirical exploration. The Quarterly Journal of Economics, 118(4), 1279-1333. https://doi.org/10.1162/003355303322552801

Campion, M. A., Fink, A. A., Ruggeberg, B. J., Carr, L., Phillips, G. M., & Odman, R. B. (2011). Doing competencies well: Best practices in competency modelling. Personnel Psychology, 64(1), 225-262. https://doi.org/10.1111/j.1744-6570.2010.01207.x

Cedefop. (2019-2024). Skills in online job advertisements. https://www.cedefop.europa.eu/en/tools/skills-intelligence/trend-focus/skills-online-job-advertisements.

Chen, H., Sun, Y., & Yuan, Y. (2026). Artificial Intelligence and Skills: Evidence from Contrastive Learning in Online Job Vacancies. General Economics. https://doi.org/10.48550/arXiv.2601.03558

Deming, D. J. (2017). The growing importance of social skills in the labor market. Quarterly Journal of Economics, 132(4), 1593-1640. https://doi.org/10.1093/qje/qjx022

Dierdorff, E. C., & Morgeson, F. P. (2007). Consensus in work role requirements: The influence of discrete occupational context on role expectations. Journal of Applied Psychology, 92(5), 1228–1241. https://doi.org/10.1037/0021-9010.92.5.1228

Frey, C. B., & Osborne, M. A. (2017). The future of employment: How susceptible are jobs to computerisation? Technological Forecasting and Social Change, 114, 254-280. https://doi.org/10.1016/j.techfore.2016.08.019

Goos, M., Manning, A., & Salomons, A. (2014). Explaining job polarization: Routine-biased technological change and offshoring. American Economic Review, 104(8), 2509-2526. https://doi.org/10.1257/aer.104.8.2509

Kohnová, L., Papula, J., Papulová, Z., Stachová, K., & Stacho, Z. (2020). Job mismatch: The phenomenon of overskilled employees as a result of poor managerial competences. Entrepreneurship and Sustainability Issues, 8(1), 83-102. http://doi.org/10.9770/jesi.2020.8.1(6)

Morgeson, F. P., & Humphrey, S. E. (2006). The Work Design Questionnaire (WDQ): Developing and validating a comprehensive measure for assessing job design and the nature of work. Journal of Applied Psychology, 91(6), 1321–1339. https://doi.org/10.1037/0021-9010.91.6.1321

National Classification of Occupations (n.d.). https://www.sustavapovolani.sk

Noe, R. A., Clarke, A. D. M., & Klein, H. J. (2014). Learning in the twenty-first-century workplace. Annual Review of Organizational Psychology and Organizational Behavior, 1, 245-275. https://doi.org/10.1146/annurev-orgpsych-031413-091321

Peterson, N. G., Mumford, M. D., Borman, W. C., Jeanneret, P. R., Fleishman, E. A., Levin, K.Y., Campion, M.A., Mayfield, M. S., Morgeson, F.P., Pearlman, K., Gowing, M.K., Lancaster, A. R., Silver, M. B., & and Dye, D. M. (2001). Understanding work using the occupational information network (O*NET): implications for practice and research. Personnel Psychology, 54(2), 451-492. https://doi.org/10.1111/j.1744-6570.2001.tb00100.x

Poláková, M., Horváthová Suleimanová, J., Madzík, P., Copuš, L., Molnárová, I., Polednová, J. (2023). Soft skills and their importance in the labour market under the conditions of Industry 5.0. Heliyon, 9(8), e18670, https://doi.org/10.1016/j.heliyon.2023.e18670

Rogiers, P., & Collings, D. G. (2024). The end of jobs? Paradoxes of job deconstruction in organizations. Academy of Management Perspectives, 38(2), 177-196. https://doi.org/10.5465/amp.2022.0236

Smerek, L., & Vetráková, M. (2020). Difference in Human Resources Development in Various Types of Companies. Polish Journal of Management Studies, 21(2), 398-411. https://doi.org/10.17512/pjms.2020.21.2.28

Teece, D. J. (2018). Business models and dynamic capabilities. Long Range Planning, 51(1), 40-49.

van Laar, E., van Deursen, A. J. A. M., van Dijk, J. A. G. M., & de Haan, J. (2017). The relation between 21st-century skills and digital skills: A systematic literature review. Computers in Human Behavior, 72, 577-588. https://doi.org/10.1016/j.chb.2017.03.010

Vuorikari, R., Kluzer, S., & Punie, Y. (2022). DigComp 2.2: The Digital Competence Framework for Citizens - With new examples of knowledge, skills and attitudes, EUR 31006 EN, Publications Office of the European Union, Luxembourg. https://doi.org/10.2760/115376, JRC128415

Zhou, H., Wang, L., Cao, Y., & Jincheng, L. (2025). The impact of artificial intelligence on labor market: A study based on bibliometric analysis. Journal of Asian Economics, 98, 101926. https://doi.org/10.1016/j.asieco.2025.101926

LIFELONG LEARNING AS A DRIVER OF SUSTAINABLE COMPETITIVENESS IN SELECTED CEE COUNTRIES

Cecília Olexová, Kristína Herzigová, Natália Olexa

Bratislava University of Economics and Business
Faculty of Business Economics with seat in Košice, Slovakia
cecilia.olexova@euba.sk, kristina.herzigova@euba.sk, natalia.olexa@student.upjs.sk

DOI: 10.35011/IDIMT-2026-383

Keywords

CEE countries; lifelong learning; digital learning; human capital; labour market.

Abstract

Current economic development is determined by the twin transition, which changes labour market structures, production models, and skill requirements of the workforce. The expansion of artificial intelligence, automation, and data-driven technologies occurs in parallel with increasing pressure for environmental sustainability. The development of digital competences and green skills represents a fundamental prerequisite for long-term sustainable competitiveness. Lifelong learning thus plays a systemic role in strengthening human capital and facilitating the transition towards knowledge- and technology-oriented economies. The aim of this paper is to analyse lifelong learning as a determinant of sustainable competitiveness in selected Central and Eastern European countries. The analysis is based on selected indicators from Eurostat and CEDEFOP, focusing on adult participation in education, online learning, employment rate, the share of employment in knowledge-intensive activities, and the number of ICT specialists, thereby enabling an assessment of the relationship between skills development and the structural transformation of economies in the era of twin transition.

1. Introduction

Ongoing shifts towards digitalisation and sustainability are transforming labour market dynamics, production systems, and the skills demanded by employers. In this context, the ability of economies to adapt increasingly depends on the development of human capital and continuous skills upgrading.

The capacity of an economy to create knowledge-intensive jobs, increase the share of ICT specialists, and support innovation activities is closely linked to the participation of the adult population in formal and non-formal education, including online learning. This highlights the key role of lifelong learning as a mechanism for enhancing employability, adaptability, and the development of digital competences.

At the same time, investments in education and skills contribute to economic performance and support the transition towards knowledge-based economies. However, significant differences persist across countries, particularly in Central and Eastern Europe (CEE), where structural transformation is still ongoing.

The aim of this paper is to analyse lifelong learning as a determinant of sustainable competitiveness in selected Central and Eastern European countries, with a focus on its relationship with digital learning and labour market outcomes.

2. Theoretical background

The concept of lifelong learning is widely recognised as a key factor in the development of skills and competences in modern economies. It refers to the continuous process of acquiring knowledge and abilities throughout an individual's life, across different contexts and forms of education (Thwe & Kálmán, 2024; García-Toledano Mayoral et al., 2023). Lifelong learning includes both formal and non-formal education and plays an essential role in improving employability and adaptability in changing labour market conditions.

Closely related to this concept is digital learning, which reflects the increasing role of technology in education. Digital learning can be understood as both a spontaneous and structured process supported by digital technologies, enabling the development of critical thinking, problem-solving, and collaborative skills (Gonçalves et al., 2017). These forms of learning are particularly important in the context of ongoing economic and technological changes, where continuous upskilling becomes necessary.

The importance of education and skills development is further emphasised by the concept of human capital, which is considered a key factor for enterprise performance and competitiveness. Research shows that investments in education not only increase work productivity within organisations but also improve individuals' socioeconomic status (Hitka et al., 2021). Similarly, Kucharčíková et al. (2023) confirm the positive impact of investments in employees' education and the implementation of human capital management on business performance, particularly in the context of Slovak enterprises. These findings highlight that lifelong and digital learning are not only beneficial at the individual level but also crucial for organisational success. While these studies primarily focus on firm-level outcomes, this paper extends the perspective to the macroeconomic level, examining how investments in education and skills are reflected in broader labour market performance and structural economic changes.

These developments are embedded in the broader context of the so-called "twin transition", which refers to the interaction between digital and green transformation. The literature emphasises that these two processes are complementary and mutually reinforcing, contributing to long-term sustainable competitiveness (Kirov, 2023; Dabbous et al., 2023; Ogrean & Herciu, 2021). Digital technologies support more efficient and sustainable economic systems, while sustainability goals drive the adoption of innovative digital solutions.

In this context, sustainability can be understood as the ability to meet present needs without compromising the ability of future generations to meet their own needs, while balancing economic, social, and environmental dimensions. The integration of digitalisation, lifelong learning, and human capital development thus creates a comprehensive framework for understanding how continuous skills development contributes not only to labour market outcomes and employability, but also to enterprise performance and long-term economic resilience.

Based on these theoretical considerations, the key concepts of lifelong learning, digital learning, and human capital are operationalised through selected indicators in the empirical analysis.

3. Data and methods

The analysis is based on panel data obtained from the Cedefop database, covering selected Central and Eastern European countries (Bulgaria, Czechia, Estonia, Croatia, Latvia, Lithuania, Hungary, Poland, Romania, Slovenia, Slovakia) over the period 2019–2024. The dataset includes harmonised indicators related to lifelong learning, digital learning, and labour market performance, ensuring cross-country comparability. The data were further aggregated into country and yearly averages for descriptive, index, and correlation analysis. The indicators include lifelong learning (1050), online learning (1140a), and employment rate (3060), which are expressed as percentages; employment in knowledge-intensive activities (3075), measured as a share of total employment; and ICT specialists (3080), measured as the number of employed individuals in thousands. These indicators were selected to capture key dimensions of lifelong learning, digital preparedness, labour market performance, and the transition towards knowledge-intensive economic activities.

The study applies a composite indicator approach to capture the multidimensional nature of digital readiness and human capital development. Since these phenomena cannot be adequately represented by a single variable, multiple indicators are aggregated into synthetic indices (OECD, 2008).

All variables were normalised using the min–max transformation, which ensures comparability across indicators with different units and scales. The indices were constructed using equal weighting and arithmetic mean aggregation, reflecting a transparent and commonly applied approach in the absence of strong theoretical justification for differential weights. This approach ensures methodological transparency and comparability across countries, which is particularly important in cross-country analyses.

Two composite indices were developed. The Digital Readiness Index (DRI) (European Commission, 2018) captures the level of digital preparedness and is based on three indicators: online learning, ICT specialists, and employment in knowledge-intensive activities. To account for potential distortions caused by the absolute nature of ICT employment, a simplified version of the index excluding this indicator was constructed. The extended version is used for the main analysis and visualisation, while the simplified version serves as a robustness check to assess the sensitivity of results.

In addition, a simplified Human Capital Index (HCI) (Kraay, 2018) was developed to reflect the relationship between skills and labour market outcomes. This index combines lifelong learning, online learning, and the employment rate, preserving the core logic that human capital is determined by both skills development and economic participation.

Finally, the relationships between selected indicators and indices were analysed using correlation analysis. The distribution of variables was assessed using graphical methods (histograms) and descriptive statistics (skewness and kurtosis), indicating approximate normality for most variables, except for ICT specialists. In addition, the Shapiro–Wilk test indicated statistically significant deviations from normality ($p < 0.05$), reflecting its sensitivity in larger samples. Therefore, Pearson correlation was used as the main analytical approach, while Spearman rank correlation was applied as a robustness check. This combined approach allows for assessing both linear relationships and the stability of the findings.

4. Results

This section presents the main trends in lifelong learning, digital learning, and labour market indicators over the study period, together with composite measures of digital readiness and human capital, and the results of the analysis of relationships between these variables.

The data presented in Table 1 indicate a general upward trend across most analysed indicators over the study period. Lifelong learning participation increased from 10.16% in 2019 to 14.03% in 2024, despite a temporary decline in 2020. A more pronounced increase can be observed in online learning, which rose significantly from 18.23% to 30.51%, with the most notable increase occurring between 2020 and 2021.

Table 1. Trends of selected indicators (2019–2024)

Study period	1050	1140a	3060	3075	3080
	Lifelong learning (%)	Online learning (%)	Employment rate (%)	Knowledge-intensive employment (%)	ICT specialists (thousands)
2019	10.16	18.23	74.12	35.23	296.13
2020	8.65	23.96	73.16	36.30	317.95
2021	11.08	29.48	74.43	37.43	336.83
2022	12.42	26.99	76.23	37.33	353.06
2023	13.11	28.56	76.75	37.37	369.63
2024	14.03	30.51	77.38	37.85	388.81

Source: authors' own processing (data: Cedefop 2019-2024)

Labour market indicators also show a positive development. The employment rate slightly decreased in 2020 but subsequently increased steadily, reaching 77.38% in 2024. Similarly, the share of employment in knowledge-intensive activities grew from 35.23% to 37.85%, indicating a gradual shift towards a knowledge-based economy.

Finally, the number of ICT specialists shows a consistent and substantial increase throughout the entire period, rising from 296.13 thousand in 2019 to 388.81 thousand in 2024. Overall, the results suggest a parallel development of digital learning and labour market performance indicators. These trends may also reflect broader labour market transformations associated with digitalisation and automation, which increase the demand for continuous upskilling. In this context, lifelong learning becomes an important mechanism supporting adaptation to changing competence requirements.

To examine the relationships between learning indicators and labour market outcomes, Pearson correlation analysis was conducted using panel data including all country-year observations over the period 2019–2024. Spearman rank correlation was additionally used to test the robustness of the results. The aim was to identify whether higher participation in lifelong and digital learning is associated with improved labour market performance and structural shifts towards a knowledge-based economy.

The results indicate a strong positive relationship between lifelong learning and online learning ($r = 0.817$), suggesting that countries with higher participation in formal and non-formal education are also associated with higher engagement in digital learning activities. A moderate positive correlation was found between lifelong learning and the employment rate ($r = 0.532$), indicating that increased participation in education is associated with better labour market outcomes. Furthermore, a similarly strong positive relationship was observed between online learning and employment in knowledge-intensive activities ($r = 0.798$), suggesting a close association between digital learning and the development of the knowledge economy. In contrast, the Pearson correlation indicates a very weak relationship between online learning and the number of ICT specialists ($r = 0.096$), suggesting that the growth of ICT employment may depend on additional structural or institutional factors beyond general digital learning participation. This finding may reflect a potential time lag between participation in digital learning and the development of specialised ICT skills, or reflect that general digital competencies do not necessarily translate into highly specialised technical occupations.

Spearman correlation confirms the robustness of the remaining relationships, as their direction and magnitude (lifelong learning and online learning: $\rho = 0.789$; lifelong learning and employment rate: $\rho = 0.561$; online learning and knowledge-intensive employment: $\rho = 0.849$) remain broadly consistent with the Pearson results. However, in the case of online learning and ICT specialists, the Spearman coefficient indicates a moderate negative association ($\rho = -0.393$). This divergence suggests that the relationship is sensitive to the distribution of the data and may reflect non-linear patterns or cross-country structural differences. This is consistent with the previously observed deviations from normality in the ICT variable. Therefore, the relationship between online learning and ICT specialists should be interpreted with caution. This finding highlights the importance of complementing linear correlation analysis with non-parametric methods when analysing heterogeneous cross-country data.

To complement the analysis of individual indicators, composite measures of digital readiness and human capital were examined (Figure 1, Figure 2). The comparison of DRI, calculated as the average of annual values over the period 2019–2024 using the extended specification, reveals significant differences in digital preparedness across countries.

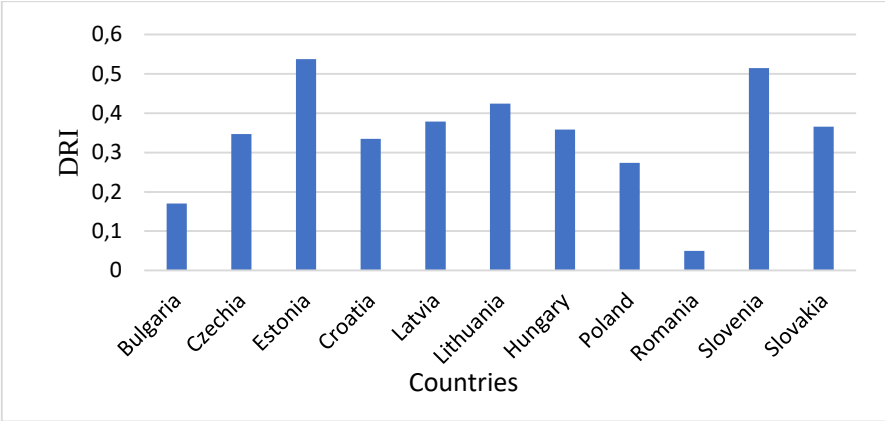


Figure 1. Average DRI by country Source: authors' own processing (data: Cedefop 2019-2024)

Estonia and Slovenia achieve the highest values (approximately 0.54 and 0.51), indicating a strong level of digital readiness. In contrast, Romania and Bulgaria show the lowest values (around 0.05 and 0.17), suggesting a substantial gap in digital development. Most countries are concentrated in the mid-range, including Lithuania, Latvia, Hungary, Slovakia, and Croatia, with values between approximately 0.33 and 0.42. This indicates a moderate level of digital readiness across the region, with noticeable disparities between leading and lagging countries.

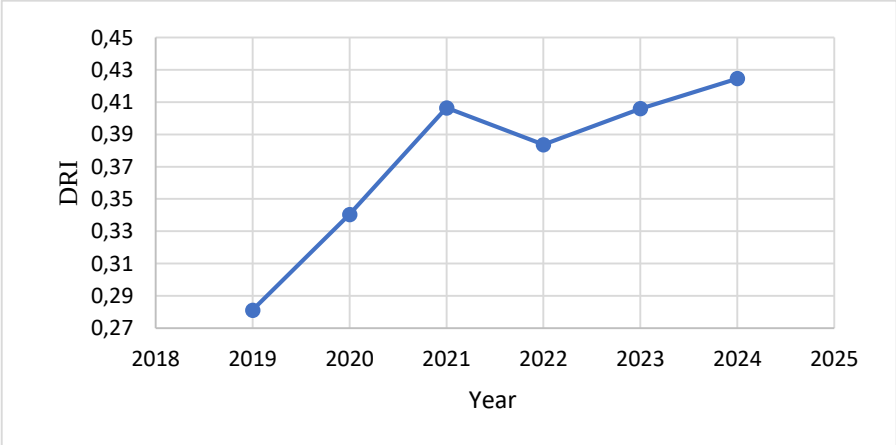


Figure 2. Development of average DRI (2019–2024) Source: authors' own processing (data: Cedefop 2019-2024)

The trend analysis, based on the average values of the extended DRI across all analysed countries, shows a general increase over the period 2019–2024. After a sharp rise between 2019 and 2021, a slight decline is observed in 2022, followed by a renewed upward trend reaching the highest value in 2024. Overall, the results indicate a gradual improvement in digital readiness over time.

HCI, calculated as the average of annual values over the period 2019–2024 for each country, reveals notable differences in the level of skills and labour market integration across the analysed countries (Figure 3).

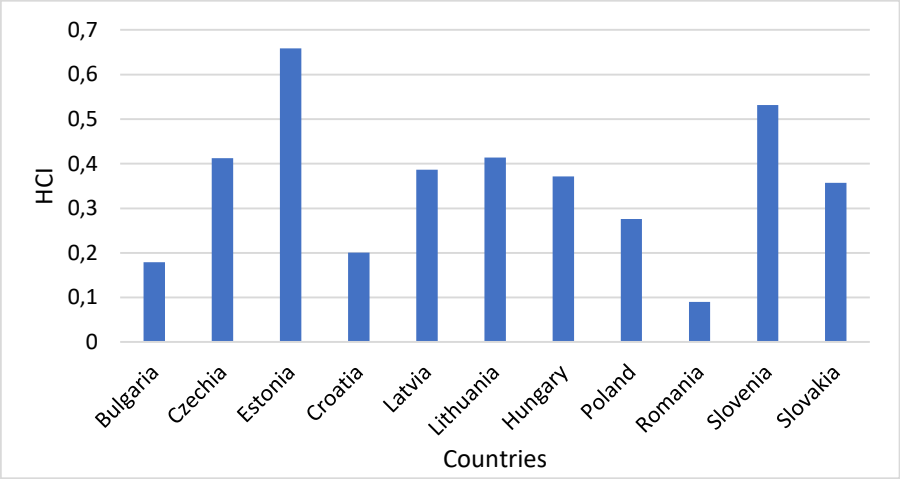


Figure 3. Average HCI by country Source: authors' own processing (data: Cedefop 2019-2024)

Estonia achieves the highest value (approximately 0.66), followed by Slovenia (around 0.53), indicating a strong performance in both skills development and employment outcomes. Countries such as Czechia, Lithuania, and Latvia reach mid-to-higher values (around 0.39–0.41), suggesting a relatively balanced level of human capital development. In contrast, Romania and Bulgaria show the lowest values (approximately 0.09 and 0.18), highlighting significant gaps in skills and labour market participation.

Overall, most countries fall within a moderate range, while the results point to clear disparities between leading and lagging economies in terms of human capital.

In addition, the relationships between composite indices and selected indicators were analysed to provide a more comprehensive view of digital readiness and human capital.

To avoid circular relationships, correlations were analysed only between indices and indicators not included in their construction; specifically, the DRI was not correlated with employment in knowledge-intensive activities, and the HCI was not correlated with the employment rate. The results indicate a moderate positive relationship between digital readiness (simplified) and employment ($r = 0.601$), while the extended index shows a weaker association ($r = 0.428$), suggesting potential distortions caused by ICT-related indicators. In contrast, HCI exhibits a strong positive correlation with employment in knowledge-intensive sectors ($r = 0.783$), suggesting an important role of skills and lifelong learning in the development of the knowledge economy.

The Spearman correlation analysis supports the Pearson results, as both the direction and the overall strength of the relationships remain largely consistent (DRI simplified and employment rate: $\rho = 0.501$; DRI extended and employment rate: $\rho = 0.414$; HCI and knowledge-intensive employment: $\rho = 0.803$). Although minor differences were observed, particularly in the case of the simplified version of DRI, these variations do not alter the substantive interpretation of the findings.

5. Conclusion

The results of the analysis suggest that lifelong learning and digital learning play an important role in shaping labour market outcomes and supporting the transition towards knowledge-based economies. The findings indicate a positive relationship between participation in education and employment performance, as well as a strong association between digital learning and the development of knowledge-intensive activities.

These results are largely consistent with the theoretical assumptions emphasised in the literature, which highlight the importance of lifelong learning and human capital development for competitiveness and economic performance (e.g. Hitka et al., 2021; Kucharcikova et al., 2023) and building of knowledge society (Antosova & Csikosova, 2013). The robustness of these findings is further supported by the Spearman correlation analysis, which confirms the consistency of the main observed relationships, while also indicating that the association between online learning and ICT specialists should be interpreted more cautiously. However, the weak relationship between online learning and the number of ICT specialists suggests that digital learning alone may not be sufficient to generate structural changes in the labour market, pointing to the relevance of additional institutional and structural factors.

At the same time, the study indicates substantial differences in digital readiness and human capital development across the analysed CEE countries, reflecting varying levels of progress in adapting to the twin transition. This suggests that policies aimed at increasing participation in lifelong and digital learning should be complemented by targeted measures supporting innovation ecosystems and labour market structures.

This also points to the growing importance of cooperation between employers, educational institutions, and policymakers in developing flexible forms of workforce education adapted to changing labour market needs. In this context, micro-credentials may represent one of the practical tools supporting more targeted and continuous skills development.

Overall, lifelong learning appears to be closely linked to the capacity of human capital to adapt to ongoing structural changes. Future research could examine how institutional conditions, education systems, and labour market policies shape its effectiveness across different national contexts.

Acknowledgement

This contribution was supported by KEGA 020EU-4/2025 “The Development of a Model for an Innovative Didactic Concept to Enhance Green Skills of Students at Economics-focused Universities with an Emphasis on Ecosystem Sustainability“.

References

Antosova, M., & Csikosova, A. (2013). Building of knowledge society through support of lifelong education. In ICETA 2013, 11th IEEE International Conference on Emerging Elearning Technologies and Applications (pp. 11-15).

Dabbous, A., Aoun Barakat, K., & Kraus, S. (2023). The impact of digitalization on entrepreneurial activity and sustainable competitiveness: A panel data analysis. Technology in Society, 73, 102224. https://doi.org/10.1016/j.techsoc.2023.102224

European Commission. (2018). Digital economy and society index (DESI) 2018: Methodological note. Publications Office of the European Union. https://ec.europa.eu/information_society/newsroom/image/document/2018-20/desi-2018-methodology_E886EDCA-B32A-AEFB-07F5911DE975477B_52297.pdf

García-Toledano Mayoral, E., Gracia-Zomeño, A., Farinho, P., & Correia Marujo Picado, L. M. (2023). Analyzing teachers' perception of the development of lifelong learning as personal, social and learning to learn competence in university students. Education Sciences, 13(11), 1086. https://doi.org/10.3390/educsci13111086

Gonçalves, A.C., Sousa, M.J., Cruz, R. (2017). Designing higher education digital course to boost entrepreneurship competencies. In Edulearn17 Proceedings. 5178-5184. https://doi.org/10.21125/edulearn.2017.2157

Hitka, M., Starchon, P., Lorincova, S., & Caha, Z. (2021). Education as a key in career building. Journal of Business Economics and Management, 22(4), 1065-1083. https://doi.org/10.3846/jbem.2021.15399

Kirov, V. (2023). Decent, inclusive, and green? Mission impossible? Social Sciences, 12(8), 433. https://doi.org/10.3390/socsci12080433

Kraay, A. (2018). Methodology for a World Bank human capital index (Policy Research Working Paper No. 8593). World Bank Group. https://thedocs.worldbank.org/en/doc/841571538503209726-0140022018/render/HCIMethodologyPaper14Sept2018.pdf

Kucharcikova, A., Miciak, M., Tokacíkova, E., & Staffenova, N. (2023). The Investments in Human Capital within the Human Capital Management and the Impact on the Enterprise's Performance. Sustainability, 15(6), 5015. https://doi.org/10.3390/su15065015

Ogrean, C., & Herciu, M. (2021). Romania's SMEs on the way to EU's twin transition to digitalization and sustainability. Studies in Business and Economics, 16(2), 282–295. https://doi.org/10.2478/sbe-2021-0040

Organisation for Economic Co-operation and Development (OECD). (2008). Handbook on constructing composite indicators: Methodology and user guide. OECD Publishing. https://doi.org/10.1787/9789264043466-en

Thwe, W.P., Kalman, A. (2024). Lifelong Learning in the Educational Setting: A Systematic Literature Review. Asia-Pacific Education Researcher, 33, 407–417. https://doi.org/10.1007/s40299-023-00738-w

EARLY CAREER & STUDENT SHOWCASE

ARTIFICIAL INTELLIGENCE ADOPTION IN SMALL AND MEDIUM-SIZED ENTERPRISES: THE MODERATING EFFECT OF FIRM SIZE IN V4 COUNTRIES

Sabina Janikovičová, František Pollák

Bratislava University of Economics and Business, Slovakia
Faculty of Business management
sabina.janikovicova@euba.sk, frantisek.pollak@euba.sk

DOI: 10.35011/IDIMT-2026-393

Keywords

Generative artificial intelligence; AI adoption; digital inequalities; digital transformation.

Abstract

The study analyses the adoption of artificial intelligence in enterprises in the V4 countries with the aim of identifying differences based on firm size and national context. It focuses on examining how the size of enterprises influences the rate of AI adoption and how this adoption has evolved over time. The analysis is based on secondary data obtained from OECD and applies descriptive and comparative statistical methods. The study covers the period 2020–2024 and distinguishes between enterprises with 10–49 and 50–249 employees. The results indicate that firm size is a key determinant of AI adoption, with medium-sized enterprises consistently achieving higher adoption rates than smaller firms. A significant increase in AI adoption is observed across all countries, particularly in the most recent period. At the same time, notable differences between countries persist, with the highest adoption rates recorded in the Czech Republic and the lowest in Poland. The findings highlight structural disparities in the adoption of artificial intelligence among enterprises and point to the importance of resources, technological readiness, and organizational capacity. The study contributes to a better understanding of digital transformation in the business sector and provides insights for policies aimed at supporting AI adoption, particularly among smaller enterprises.

1. Introduction

The development of digital technologies in recent years has fundamentally changed the way the economy and business environment operate. One of the most significant trends today is the increasing adoption of artificial intelligence, which represents a key driver of digital transformation and innovation. AI technologies enable enterprises to improve efficiency, optimize processes, and enhance decision-making, thereby strengthening their competitiveness in a rapidly evolving market environment (Dorcak et al., 2014; Dorcak et al., 2015). As a result, the integration of artificial intelligence into business operations has become an important strategic priority for many organizations (Jin et al., 2025; Aldreabi et al., 2025). Despite the growing importance of AI, its adoption is not evenly distributed across enterprises. Existing research indicates that the

implementation of advanced technologies is influenced by several organizational factors, among which firm size plays a crucial role. Larger and medium-sized enterprises typically have greater access to financial, technological, and human resources, which allows them to adopt and implement AI solutions more effectively. In contrast, smaller enterprises often face barriers such as limited resources, lack of expertise, limited digital readiness, integration capabilities, and shortage of specialised skills, which may slow down their adoption of artificial intelligence (Weilbach, 2025; Muhammad et al., 2025). In addition to organizational factors, differences in AI adoption can also be observed at the national level. Countries vary in terms of digital infrastructure, innovation capacity, and policy support, which significantly affects the ability of enterprises to adopt new technologies. The Visegrad Group countries – Czech Republic, Slovakia, Poland and Hungary – provide a suitable context for analysing these differences. Although these countries share similar historical and economic backgrounds, they differ in the pace and intensity of digital transformation (Belas et al., 2025; Záhorec & Hašková, 2025). The aim of this paper is to analyse the adoption of artificial intelligence in enterprises in the V4 countries, with a specific focus on the role of firm size. The study examines differences between small enterprises (10–49 employees) and medium-sized enterprises (50–249 employees) and evaluates the development of AI adoption over the period 2020–2024. The paper seeks to contribute to a better understanding of the factors influencing AI adoption in the business sector and to highlight the implications for digital transformation and competitiveness of enterprises. Despite the growing body of literature on artificial intelligence adoption, limited research has comparatively examined AI adoption patterns across the Visegrad Group countries while explicitly considering firm size as a structural factor influencing adoption. Existing studies often focus either on single-country contexts or on broader international comparisons without addressing the specific regional dynamics of Central and Eastern Europe. This creates a research gap regarding how enterprise size interacts with national context in shaping AI adoption. This study addresses this gap by analysing AI adoption in enterprises across Czechia, Slovakia, Poland, and Hungary, with a specific focus on differences between small (10–49 employees) and medium-sized enterprises (50–249 employees). The findings contribute to a better understanding of digital inequalities among SMEs and provide practical implications for policymakers designing targeted digital transformation.

2. Literature review

The way in which businesses adopt digital technologies, including artificial intelligence, has long been a topic of research within the framework of theories of technology diffusion and technology acceptance. According to the available literature, the most influential approaches include the Technology Acceptance Model, which emphasizes perceived usefulness and ease of use as key determinants of technology adoption. Extensions of this model incorporate not only individual factors but also social and organizational influences, thereby providing a more comprehensive view of technology adoption in business environments (Marikyan & Papagiannidis, 2024; Mogaji et al., 2024; Davis et al., 2024). In the context of artificial intelligence, recent research has increasingly focused on its application within organizations and its role in enhancing business performance and competitiveness. Although generative AI tools represent a rapidly evolving segment of AI technologies, their adoption in enterprises is shaped not only by individual capabilities but also, more importantly, by organizational characteristics and strategic priorities. Existing studies suggest that the adoption of AI is influenced by a combination of organizational and macroeconomic factors, including technological readiness, managerial support, and access to resources (Kshetri et al., 2024; Rodriguez Guzman, 2024; Donvir et al., 2024; Grewal et al., 2025). Artificial intelligence represents a broad category of technologies enabling systems to perform tasks that typically require human cognitive abilities, including machine learning, computer vision, robotics, expert systems, and natural

language processing. Large language models (LLMs), including generative AI tools, represent a narrower subset of AI technologies specifically focused on language generation and understanding. Since the OECD indicator analysed in this study measures enterprise adoption of artificial intelligence in general rather than specific use of generative AI or LLMs, the analysis refers to AI adoption in the broader sense (Grewal et al., 2025). At the organizational level, firm size emerges as one of the most significant determinants of AI adoption. Larger and medium-sized enterprises generally have better access to financial, human, and technological resources, which enables them to implement advanced technologies more effectively (Badghish & Soomro, 2024; Janzen, 2025). In contrast, smaller enterprises, particularly SMEs, often face limitations related to insufficient capital, lack of expertise, and underdeveloped digital infrastructure, which may hinder the adoption of AI solutions. At the same time, the literature emphasizes that AI adoption in enterprises is an integral part of a broader digital transformation process, involving changes in organizational processes, business models, and strategic management (Almashawreh et al., 2024; Abrokwah-Larbi & Awuku-Larbi, 2024; Olukoya et al., 2025). From a macroeconomic perspective, national and regional differences also play an important role in shaping the rate of AI adoption in businesses. Factors such as the level of economic development, quality of digital infrastructure, innovation capacity, and public policies significantly influence the ability of enterprises to adopt and implement AI technologies. In this regard, the countries of the Visegrad Group are often analysed as a relatively homogeneous region with similar historical and economic backgrounds, yet with noticeable differences in the pace and intensity of digital transformation (Pollák et al., 2023; Bednarzewska & Zinczuk, 2024; Prochádzka, 2025; Hruby, 2025). Recent research also highlights the growing importance of AI technologies, including generative AI, in transforming business operations and decision-making processes. These technologies have the potential to lower barriers to entry for advanced digital solutions and support innovation across firms of different sizes. However, despite these opportunities, disparities in adoption persist, particularly between firms of different sizes and levels of technological readiness. Therefore, AI adoption can be understood as a multidimensional phenomenon shaped by organizational capabilities and the broader institutional environment (Bencsik, 2022; Kaniok & Hloušek, 2025; Fedorko & Kizák, 2024; Misiak-Kwit et al., 2025).

3. Methodology

The aim of this study is to analyse the adoption of artificial intelligence in enterprises within the Visegrad Group (V4) countries, with a specific focus on the role of firm size as a factor associated with differences in adoption intensity. The analysis is based on secondary data obtained from the OECD Going Digital Toolkit database, specifically the indicator "Businesses using artificial intelligence". This indicator measures the percentage of enterprises reporting the use of at least one artificial intelligence technology in business operations. The OECD compiles these data from harmonised national ICT usage surveys conducted among enterprises, which ensures international comparability across countries. The study covers the period 2020–2024 and focuses on four countries: Czechia, Slovakia, Poland, and Hungary. At the time of analysis, 2025 data were not consistently available for all analysed countries; therefore, the study focuses on the 2020–2024 period to preserve comparability. The analysed indicator is disaggregated by enterprise size: a) enterprises with 10–49 employees, b) enterprises with 50–249 employees.

The methodological procedure consisted of four steps:

1. Extraction of OECD data for the selected V4 countries for the period 2020–2024.
2. Classification of enterprises according to firm size categories.
3. Cross-country comparison of AI adoption rates within each size category.
4. Longitudinal analysis of changes in AI adoption over time.

The study applies descriptive and comparative statistical methods. Since the dataset consists of aggregated secondary indicators, inferential statistical methods were not applied. The analysis focuses on identifying differences in adoption rates between countries and firm size categories, as well as examining temporal development patterns.

4. Results and discussion

Data on the use of artificial intelligence in companies by size in the period from 2020 to 2024 point to a gradual increase in the adoption of AI in all monitored countries, with significant differences not only between countries, but also between individual company size categories.

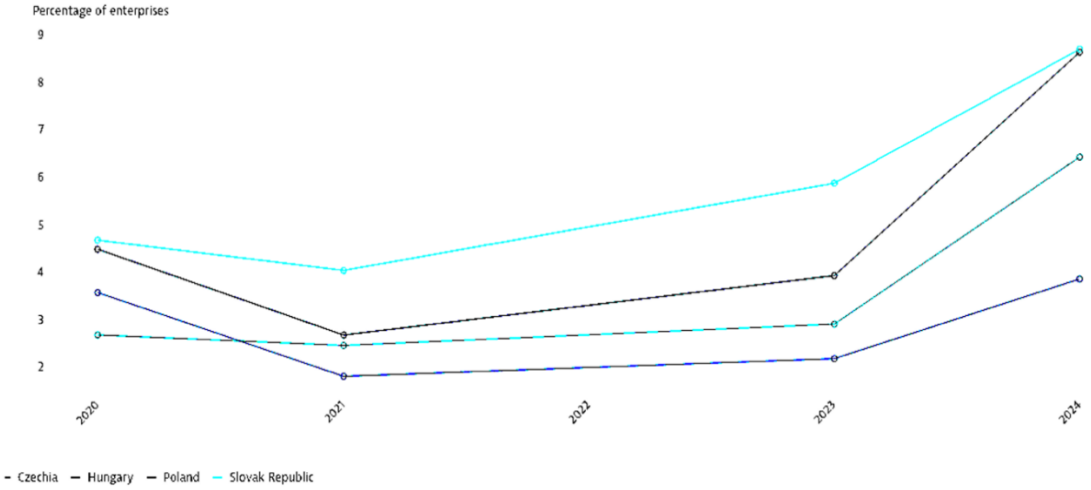


Figure 1. Share of enterprises using AI (%), firms with 10–49 employees Source: (OECD, 2025)

As we can see from the following table:

Table 1. Share of enterprises using artificial intelligence (% of all enterprises with 10–49 employees)

Year	Czechia	Slovakia	Poland	Hungary
2020	4.56	4.75	3.64	2.75
2021	2.75	4.11	1.88	2.53
2022	3.37	5.03	2.06	2.75
2023	4.00	5.95	2.25	2.98
2024	8.72	8.78	3.93	6.50

Source: own processing based on (OECD, 2025)

in the case of small and medium-sized enterprises with 10 to 49 employees, relatively low but gradually increasing values can be observed throughout the entire monitored period. In 2020, the shares of enterprises using AI reached approximately 4–5% in the Czech Republic (4.56%) and Slovakia (4.75%), while Poland (3.64%) and especially Hungary (2.75%) lagged. In 2021, there was a slight decrease in some countries, especially the Czech Republic and Poland, which may be related to the economic impacts of the previous period. However, there has been a gradual increase since 2022, which significantly accelerates in 2024. This year, Slovakia (8.78%) and the Czech Republic (8.72%) reach the highest values, while Hungary (6.50%) records a significant increase and Poland (3.93%) remains at the lowest level.

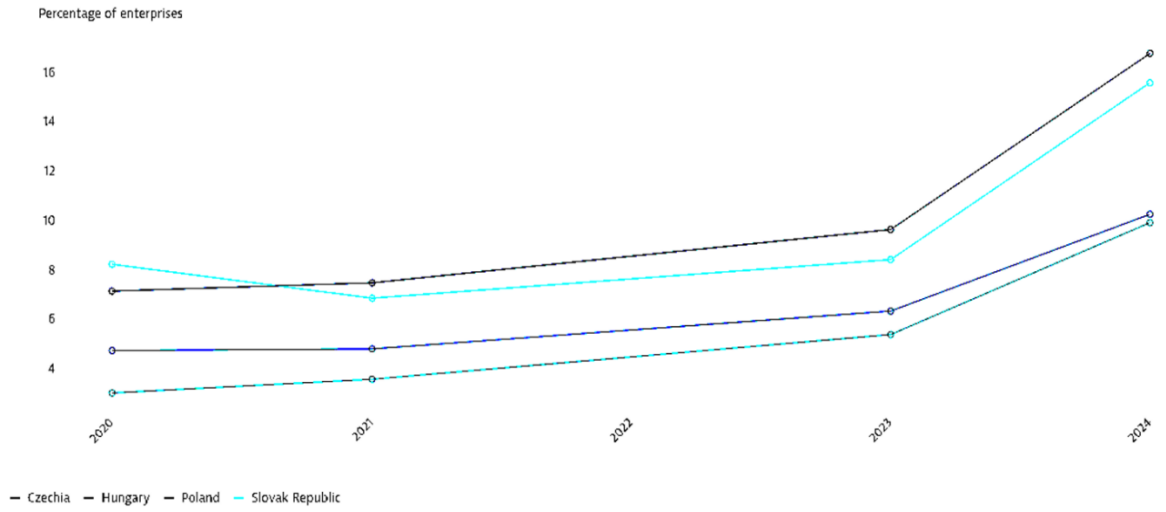


Figure 2. Share of enterprises using AI (%), firms with 50–249 employees Source: (OECD, 2025)

As we can see from the following table:

Table 2. Share of enterprises using artificial intelligence (% of all enterprises with 50–249 employees)

Year	Czechia	Slovakia	Poland	Hungary
2020	7.30	8.38	4.89	3.17
2021	7.63	7.01	4.96	3.73
2022	-	-	-	-
2023	9.79	8.57	6.49	5.54
2024	16.91	15.73	10.41	10.07

Source: own processing based on (OECD, 2025)

In the category of medium-sized enterprises with 50 to 249 employees, the values are significantly higher, indicating a strong influence of company size on the adoption rate of artificial intelligence. In 2020, Slovakia achieved the highest share (8.38%), followed by the Czech Republic (7.30%), Poland (4.89%) and Hungary (3.17%). In 2021, the values change slightly, with the Czech Republic recording an increase (7.63%), while Slovakia a slight decrease (7.01%). Data for 2022 is not available, but in 2023, more significant growth can already be observed in all countries. The highest values are achieved by the Czech Republic (9.79%), followed by Slovakia (8.57%), Poland (6.49%) and Hungary (5.54%). However, the most significant increase occurs in 2024, when the shares of enterprises using AI almost double. The Czech Republic achieves the highest value (16.91%), followed by Slovakia (15.73%), Poland (10.41%) and Hungary (10.07%). Based on the above data, several key trends can be identified. First, the adoption of artificial intelligence in enterprises is clearly growing, with the most significant increase occurring in the last monitored year, 2024. This development is likely related to the dynamic development of generative AI and the increased pressure on the digital transformation of enterprises. Second, the size of the enterprise appears to be a fundamental factor influencing the rate of use of AI. Medium-sized enterprises (50–249 employees) achieve systematically higher values than smaller enterprises (10–49 employees), which can be explained by greater financial and technological resources, as well as a higher degree of process formalization and strategic management. Third, there are also significant differences between countries. The Czech Republic and Slovakia have long been among the leaders in AI adoption in both company size categories, while Poland shows the lowest values throughout the period. Hungary is located between these groups, but in 2024 it will see significant growth, approaching the level of developed countries in the region. The observed differences between enterprise size categories may

be explained by structural disparities in organisational resources. Medium-sized enterprises typically have greater financial capacity, more formalised innovation processes, higher managerial capabilities, and better access to specialised employees, which may facilitate the adoption and implementation of AI technologies. At the national level, differences between V4 countries may reflect variations in digital maturity, innovation ecosystems, public support for digitalisation, and broader technological readiness. The stronger performance of Czechia and Slovakia may be associated with relatively more advanced digital transformation among enterprises, while the lower values observed in Poland may indicate slower diffusion of AI technologies among SMEs. From a practical perspective, the lower adoption rates among smaller enterprises suggest the need for targeted policy interventions. These may include digital training programmes, AI-focused advisory support, subsidised access to cloud-based solutions, innovation vouchers, and programmes aimed at strengthening managerial readiness.

5. Conclusion

Based on the conducted analysis, it can be concluded that the defined objective was successfully fulfilled, as the results clearly demonstrated significant differences in the use of artificial intelligence depending on the size of enterprises as well as between individual countries. The findings suggest that firm size is strongly associated with differences in AI adoption rates across enterprises. Medium-sized enterprises (50–249 employees) consistently achieved higher adoption rates compared to smaller enterprises (10–49 employees), which may reflect broader differences in organisational resources, including financial capacity, technological readiness, and human capital. At the same time, a significant increase in the adoption of artificial intelligence was observed across all analysed countries, particularly in the most recent period, highlighting the growing importance of AI as a driver of digital transformation in the business sector. Differences between countries were also identified, with the Czech Republic achieving the highest adoption rates and Poland the lowest throughout the observed period. The most important finding of this study is that the adoption of artificial intelligence in enterprises is not uniform but strongly conditioned by firm size. This implies that smaller enterprises may face structural disadvantages in adopting advanced technologies, which could affect their long-term competitiveness. Therefore, firm size should be considered a critical factor when designing strategies and policies aimed at supporting digital transformation. Despite its contributions, the study has several limitations. Firstly, it is based on secondary data obtained from OECD, which does not allow for a deeper examination of the underlying causes of the observed differences or internal processes within enterprises. Secondly, the analysis is limited to the V4 countries, which restricts the generalizability of the findings to a broader European or global context. Another limitation is the availability of data, particularly missing values for certain years (e.g. 2022), as well as the fact that the analysed indicator captures only the presence of AI adoption and does not reflect its intensity, scope, or effectiveness. Based on the findings, several directions for future research can be proposed. The study excludes micro-enterprises with fewer than 10 employees, although this category represents a substantial part of the SME sector and may face even stronger barriers to AI adoption. Future research should therefore include micro-enterprises to provide a more comprehensive understanding of digital inequalities across firm sizes. Another limitation is the absence of sectoral disaggregation in the OECD dataset. AI adoption likely differs significantly across sectors such as manufacturing, services, agriculture, and information technology. Future studies should therefore examine industry-specific adoption patterns.

Acknowledgement

This contribution is a partial output of the VEGA project no. 1/0110/24 and EUBA project no. I-26-103-00.

References

Abrokwah-Larbi, K., & Awuku-Larbi, Y. (2024). The impact of artificial intelligence in marketing on the performance of business organizations: Evidence from SMEs in an emerging economy. Journal of Entrepreneurship in Emerging Economies, 16(4), 1090–1117.

Aldreabi, H., Dahdoul, N. K. S., Alhur, M., Alzboun, N., & Alsalhi, N. R. (2025). Determinants of student adoption of generative AI in higher education. Electronic Journal of e-Learning, 23(1), 15–33.

Almashawreh, R. E., Talukder, M., Charath, S. K., & Khan, M. I. (2024). AI adoption in Jordanian SMEs: The influence of technological and organizational orientations. Global Business Review.

Aruleba, K., Jere, N., & Matarirano, O. (2022). An evaluation of technology adoption during remote teaching and learning at tertiary institution by gender. IEEE Transactions on Computational Social Systems, 10(3), 1335–1346.

Badghish, S., & Soomro, Y. A. (2024). Artificial intelligence adoption by SMEs to achieve sustainable business performance: Application of the technology–organization–environment framework. Sustainability, 16(5), 1864.

Bednarzewska, K., & Zinczuk, B. (2024). Analysis of the potential of the Visegrad Group countries in selected areas of their activities. Zeszyty Naukowe. Organizacja i Zarządzanie/Politechnika Śląska, 191, 63–75.

Belas, J., Kliestik, T., Dvorsky, J., & Streimikiene, D. (2025). Exploring gender-based disparities in the digital transformation and sustainable development of SMEs in V4 countries. Journal of Innovation & Knowledge, 10(2), Article 100681.

Bencsik, A. (2022). Artificial intelligence in the Middle East European countries. In J. M. Munoz & A. Maurya (Eds.), International perspectives on artificial intelligence (pp. 52–62). Anthem Press.

Davis, F. D., Granić, A., & Marangunić, N. (2024). The technology acceptance model: 30 years of TAM (pp. 20–54). Springer International Publishing.

Dorcak, P., Pollák, F., & Szabo, S. (2014). Analysis of the possibilities of improving an online reputation of public institutions. In Proceedings of the 22nd Interdisciplinary Information Management Talks: IDIMT-2014 Networking Societies—Cooperation and Conflict (pp. 275–282). Poděbrady, Czech Republic.

Dorcak, P., Strach, P., & Pollak, F. (2015). Analytical view of the perception of selected innovative approaches in marketing communications. Quality Innovation Prosperity, 19(1), 74–84.

Donvir, A., Panyam, S., Paliwal, G., & Gujar, P. (2024). The role of generative AI tools in application development: A comprehensive review of current technologies and practices. In 2024 International Conference on Engineering Management of Communication and Technology (EMCTECH) (pp. 1–9). IEEE.

OECD. (2025). Going Digital Toolkit: Businesses using artificial intelligence. https://goingdigital.oecd.org

Fedorko, R., & Kizák, M. (2024). In the Visegrad Group: Insights from the digital economy and society. In Eurasian Business and Economics Perspectives: Proceedings of the 43rd Eurasia Business and Economics Society Conference (p. 175). Springer Nature.

Grewal, D., Satornino, C. B., Davenport, T., & Guha, A. (2025). How generative AI is shaping the future of marketing. Journal of the Academy of Marketing Science, 53(3), 702–722.

Hruby, V. (2025). Entrepreneurial orientation in the age of artificial intelligence: A study of SMEs in the Visegrad chemical sector. Journal of East European Management Studies, 30(3), 45710.

Iddrisu, H. M., Iddrisu, S. A., & Aminu, B. (2025). Gender differences in the adoption, usage, and perceived effectiveness of AI writing tools: A study among university for development studies students. International Journal of Educational Innovation and Research, 4(1), 110–111.

Janzen, K. A. (2025). Inside AI transition: Organizational change in the making: A study of early-stage AI implementation in transitioning SMEs [Master's thesis, Malmö University]. DiVA Portal. https://mau.diva-portal.org/smash/get/diva2:1961813/FULLTEXT01.pdf.

Jin, L., Shen, Z., Alhur, A. A., & Naeem, S. B. (2025). Exploring the determinants and effects of artificial intelligence (AI) hallucination exposure on generative AI adoption in healthcare. Information Development.

Kaniok, P., & Hloušek, V. (2025). Visegrad Four as an institution in times of EU crises. European Politics and Society, 26(5), 1070–1088.

Kshetri, N., Dwivedi, Y. K., Davenport, T. H., & Panteli, N. (2024). Generative artificial intelligence in marketing: Applications, opportunities, challenges, and research agenda. International Journal of Information Management, 75, Article 102716.

Marikyan, D., & Papagiannidis, S. (2024). Technology acceptance model: A review. In S. Papagiannidis (Ed.), TheoryHub Book. Newcastle University. https://open.ncl.ac.uk/theories/1/technology-acceptance-model/

Misiak-Kwit, S., Smerek, L., Szabó-Szentgróti, G., & Marková, H. (2025). HR digitisation before and during COVID-19 in Visegrad Group countries. In S. Al-Augby, S. Alwajidi, P. Bajdor, M. Borawski, & J. Błażejewska-Jaśkowiak (Eds.), Intelligent management and artificial intelligence: Trends, challenges, and opportunities: Proceedings of the 28th European Conference on Artificial Intelligence (ECAI 2025) (Vol. 1, pp. 333–350). Uniwersytet Szczeciński.

Mogaji, E., Viglia, G., Srivastava, P., & Dwivedi, Y. K. (2024). Is it the end of the technology acceptance model in the era of generative artificial intelligence? International Journal of Contemporary Hospitality Management, 36(10), 3324–3339.

Muhammad, S. S., Dey, B. L., Kamal, M. M., Samuel, L., & Alzeiby, E. A. (2025). Digital transformation or digital divide? SMEs' use of AI during global crisis. Technological Forecasting and Social Change, 217, Article 124184.

Olukoya, D., Alumona, P., Okoro, S. C., Ibiyeye, A. O., Ibiyeye, S. O., & Ikhifa, M. O. (2025). AI in SMEs: A holistic approach for successful implementations. International Journal of Science and Research Archive, 10.

Pollák, F., & Markovič, P. (2021). Economic activity as a determinant for customer adoption of social media marketing. Sustainability, 13(7), 3999.

Pollák, F., Markovič, P., & Majdúchová, H. (2023). Reputation of electric vehicles in the environment of carbon reduction and accelerated digitization. Energies, 16(9), 3836.

Procházka, G. (2025). Artificial intelligence and sustainable development: Policy approaches from Visegrad Group countries. https://www.scilit.com/publications/068f738cd087ea4c017eec98271a91e6

Rani, V., & Kumar, J. (2024). Gender differences in FinTech adoption: What do we know, and what do we need to know? Journal of Modelling in Management, 19(4), 1215–1236.

Rodriguez Guzman, M. M. (2024). Using generative AI tools to improve marketing communication: Opportunities, challenges and best practices.

Tufa, A. H., Alene, A. D., Cole, S. M., Manda, J., Feleke, S., Abdoulaye, T., & Manyong, V. (2022). Gender differences in technology adoption and agricultural productivity: Evidence from Malawi. World Development, 159, Article 106027.

Weilbach, N. (2025). Bridging the digital divide: AI adoption for SMME sustainability in resource-constrained regions. Open Journal of Business and Management, 13(2), 1289–1300.

Záhorec, J., & Hašková, A. (2025). Comparison of digital skills upgrading of students and teachers in the digital transformation of East-Central Europe. European Journal of Contemporary Education, 14(2), 216–227.

AIROF: A PRACTICAL FRAMEWORK FOR ADAPTIVE RISK AND OPPORTUNITY MANAGEMENT IN THE USE OF AI TOOLS IN CONSULTING SMES

Ludmila Jiříčková

Prague University of Economics and Business
Faculty of Informatics and Statistics
ludmila.malinova@vse.cz

DOI: 10.35011/IDIMT-2026-401

Keywords

Artificial intelligence governance; risk management; opportunity management; SME consulting organisations; scorecard; design science research.

Abstract

Artificial intelligence (AI) is increasingly becoming an integral part of everyday organisational practice, particularly in consulting environments where knowledge work, decision-making and communication play a central role. While large organisations gradually implement structured approaches to AI governance, small and medium-sized enterprises (SMEs) often lack a practical and operationally usable framework for managing both risks and opportunities associated with AI. Existing approaches, including regulatory frameworks and management standards, provide valuable guidance but remain either too abstract or too complex for everyday managerial application in SME settings. This paper presents work-in-progress doctoral research focused on the design of AIROF (Artificial Intelligence Risk & Opportunity Framework), an adaptive framework grounded in fourth-order cybernetics, AI governance principles, risk management and the structural logic of the Balanced Scorecard. The framework is further operationalised through the AIROF Scorecard, a practical artefact designed for implementation in Microsoft Excel/VBA and structured around role-based organisational use. The intended contribution of the paper is to develop a lightweight and adaptive artefact that may help to translate selected high-level AI governance and risk management principles into more operational decision-making routines in consulting SMEs. Since the research is still in progress, the paper does not present empirical results, but rather the theoretical and methodological grounding of the proposed framework and its planned validation.

1. Introduction

The rapid development of artificial intelligence is fundamentally transforming organisational decision-making processes. AI tools are no longer limited to experimental use but are actively applied in analysis, communication, content generation and strategic decision support (Russell & Norvig, 2021; Davenport & Mittal, 2023). This transformation creates both opportunities and risks. While AI enables increased efficiency and innovation, it also introduces uncertainty related to bias, transparency, overreliance and accountability.

These challenges are particularly significant in consulting SMEs. Such organisations operate in dynamic environments, rely heavily on human expertise and maintain close relationships with clients. AI therefore directly affects their core value creation processes. At the same time, SMEs often lack structured approaches to managing AI-related risks and opportunities.

Existing governance frameworks, including the European Union Artificial Intelligence Act, the NIST Artificial Intelligence Risk Management Framework and ISO/IEC 42001, provide important reference points for AI regulation, risk management and AI management systems (European Union, 2024; NIST, 2023; ISO, 2023b). However, these frameworks primarily define principles, obligations, risk categories or management system requirements. For consulting SMEs, the remaining challenge is how to translate such guidance into simple, repeatable and role-based managerial routines. This paper therefore does not claim to replace these frameworks. It focuses on the design of a practical artefact that is intended to support everyday identification, prioritisation and discussion of AI-related risks and opportunities in consulting SMEs.

This paper addresses this gap through the design of AIROF, an adaptive framework for managing risks and opportunities associated with AI use in consulting SMEs.

2. Theoretical foundations

The conceptual foundation of AIROF builds on the integration of several theoretical domains that are typically addressed separately. A key starting point is the understanding of organisations as complex socio-technical systems, where decision-making emerges from interactions between people, technologies and context (Capra & Luisi, 2014).

The first domain is AI governance, which has emerged as a response to the increasing complexity and societal impact of artificial intelligence. International frameworks such as Organisation for Economic Co-operation and Development (OECD) principles and United Nations Educational, Scientific and Cultural Organization (UNESCO) recommendations emphasise human-centred, transparent and accountable AI systems (OECD, 2024; UNESCO, 2021).

The second domain is risk management, where frameworks such as ISO 31000 and Committee of Sponsoring Organizations of the Treadway Commission Enterprise Risk Management (COSO ERM) provide structured approaches to identifying, analysing and managing uncertainty in organisational contexts (ISO, 2018; COSO, 2017).

The third domain is the Balanced Scorecard, which offers a multi-perspective approach to linking strategy and operational decision-making (Kaplan & Norton, 1992).

The fourth and central domain is fourth-order cybernetics, which serves as the meta-framework of this research. The concept of cybernetics, originally defined by Wiener (1948) as the study of control and communication in systems, provides an important foundation for understanding feedback and adaptation mechanisms. Unlike earlier cybernetic approaches, fourth-order cybernetics emphasises reflexivity, contextual integration and adaptive learning in complex socio-technical systems. It conceptualises organisations as evolving systems that continuously interact with their environment and adjust their internal structures accordingly (Božičnik & Mulej, 2011; Medvedeva & Umpleby, 2024).

This perspective is particularly relevant in the context of AI, where systems are not only technical artefacts but integral parts of organisational decision-making processes. The key differentiator is the use of fourth-order cybernetics, which emphasises reflexivity and the role of the observer (von Foerster, 1981), contextual adaptation and system integration (Medvedeva & Umpleby, 2024), as

well as dynamic feedback and the co-evolution of the system and its environment. This perspective shifts AI governance from static compliance toward adaptive organisational capability.

3. Current research

At the current stage of the research, the theoretical background has been completed and the research methodology has been fully defined. The work is now focused on the finalisation of the AIROF conceptual framework. The next step consists of developing its operationalised form, AIROF Scorecard, implemented as a practical tool in Microsoft Excel with Visual Basic for Applications (VBA) support. This artefact will subsequently be deployed in consulting SMEs, where it will be tested and validated in real organisational environments. The validation phase will focus on usability, clarity and the contribution of the framework to decision-making processes related to AI risks and opportunities.

It is important to emphasise that this paper reports a work-in-progress stage of the doctoral research. At present, the theoretical background, research problem, research questions and methodological design have been completed. The AIROF conceptual framework is being finalised and its operational form, AIROF Scorecard, is being designed as a prototype. The empirical validation has not yet been conducted and no empirical findings are reported in this paper. The planned validation in consulting SMEs will be used to assess whether the proposed artefact is understandable, usable and practically relevant for managerial decision-making.

3.1. Problem definition

The core problem addressed in this research can be formulated as follows:

There is no practical, adaptive framework that enables SME consulting organisations to systematically manage both risks and opportunities associated with AI in everyday decision-making.

Existing frameworks define principles and structures, but they do not provide operational tools for prioritisation, integration of risk and opportunity perspectives, or sufficient adaptability to rapidly changing AI environments. This gap is particularly critical in SME contexts, where resources are limited and decision-making must be efficient and flexible.

3.2. Design science research

The research follows a Design Science Research (DSR) approach, which focuses on the design and evaluation of artefacts addressing real-world problems (Hevner et al., 2004; Peffers et al., 2007).

The process includes:

- problem identification,
- definition of objectives,
- design and development of artefacts,
- demonstration and evaluation,
- communication of results.

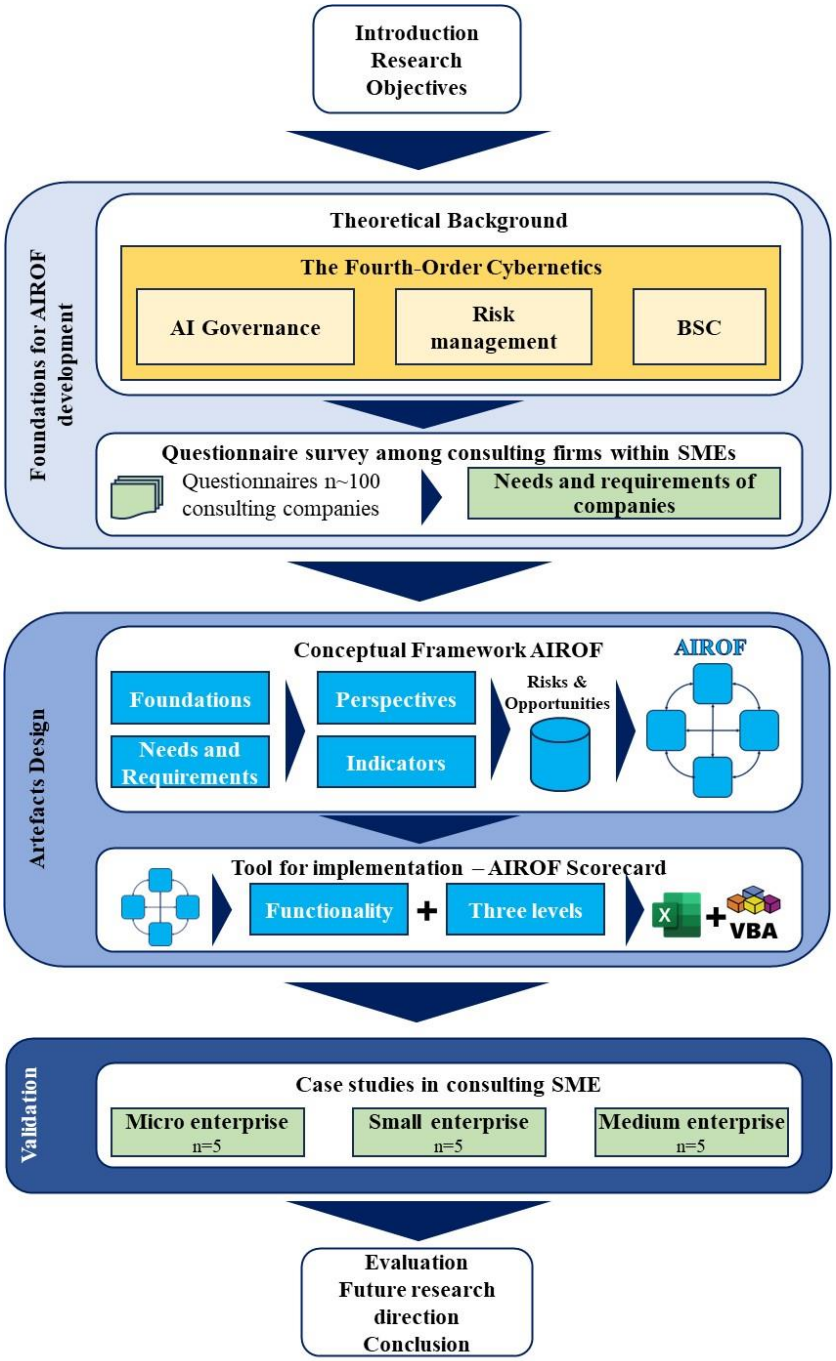


Figure 1. The application of DSR in this research Source: author

The primary objective is to design and validate a framework that enables adaptive management of AI-related risks and opportunities in consulting SMEs. The research process includes problem identification, conceptual framework development, artefact creation, implementation and evaluation.

3.3. Research questions

The main research question is:

- RQ0: How can a conceptual framework (AIROF) for the adaptive management of AI-related risks and opportunities in consulting SMEs be designed and empirically validated based on the principles of fourth-order cybernetics, AI governance and risk management?

Sub-questions include:

- RQ1: What theoretical concepts and principles of fourth-order cybernetics, AI governance, risk management and the Balanced Scorecard are relevant for designing an adaptive framework for managing AI-related risks and opportunities?
- RQ2: What are the key needs, expectations and challenges associated with the use of AI tools in consulting SMEs?
- RQ3: How can the AIROF Scorecard be designed as a practical artefact based on the identified theoretical foundations and empirical insights, while reflecting the structure of the Balanced Scorecard?
- RQ4: What is the impact of implementing the AIROF Scorecard on managing AI-related risks and opportunities in consulting organisations, particularly in terms of decision-making support, transparency and effectiveness?

The defined research questions correspond to the individual phases of the Design Science Research (DSR) process and collectively guide the design, development and evaluation of the proposed artefacts. They reflect the progression from theoretical grounding and problem identification, through artefact design and operationalisation, to empirical validation in organisational settings. In this way, the research questions not only structure the research process but also serve as a mechanism for verifying the functionality and practical relevance of the AIROF framework and the AIROF Scorecard.

3.4. Artefacts

The research produces two main artefacts:

AIROF framework

AIROF is conceptualised as an adaptive framework integrating cybernetic feedback and adaptation, AI governance principles, and the management of risks and opportunities. The framework emphasises continuous feedback loops, the integration of risk and opportunity perspectives, and adaptability to the organisational context.

AIROF scorecard

The framework is operationalised through the AIROF Scorecard, which serves as a practical tool for organisations. The scorecard enables structured recording of AI use cases, identification of risks and opportunities, prioritisation and decision-making, as well as the assignment of responsibility. This operationalisation reflects the need to translate abstract governance principles into actionable organisational practice.

The operational logic of AIROF is based on a simple iterative cycle. First, an organisation records a specific AI use case, for example the use of generative AI for analysis, communication or client deliverables. Second, the use case is assessed from both risk and opportunity perspectives. Risks may include confidentiality, bias, overreliance, lack of transparency or unclear accountability, while opportunities may include time savings, better knowledge structuring, improved service quality or innovation potential. Third, the identified risks and opportunities are prioritised through selected indicators in the AIROF Scorecard. Fourth, responsibility is assigned to a relevant organisational role and an appropriate managerial decision is recorded. Finally, the use case is reviewed again after practical experience, which creates a feedback loop and supports adaptation of the framework to the organisational context.

A simple example may be a consulting SME using generative AI to prepare a first draft of a client market analysis. In the AIROF Scorecard, this use case would be recorded together with its purpose, responsible role and expected benefit. The opportunity side may include faster preparation of background material and better structuring of information. The risk side may include inaccurate outputs, unsupported claims, use of confidential client data and insufficient human review. The managerial decision may therefore be formulated as: “allowed under conditions”, for example only with anonymised inputs, documented human verification and final approval by a responsible consultant. This example illustrates that AIROF is not intended to automate governance, but to make AI-related decisions more explicit, traceable and revisable.

3.5. Validation approach

The validation is structured into three phases. The first phase consists of a pilot implementation in three consulting SMEs representing different organisational sizes (micro, small and medium enterprises). This phase focuses primarily on the usability of the AIROF Scorecard and its initial impact on identifying and structuring AI-related risks and opportunities. Data collection is based on post-use questionnaires and semi-structured interviews with key users and managers.

The participating companies have not yet been selected at the current stage of the research. They will be recruited through purposive sampling from consulting SMEs, with the aim of covering micro, small and medium-sized organisations. The recruitment will rely on the author’s professional and academic networks and targeted contact with consulting firms that already use AI tools in knowledge work. The sample is therefore not intended to be statistically representative, but suitable for design-oriented validation of the proposed artefact.

The second phase extends the validation to a broader sample of organisations, reaching a total of fifteen consulting SMEs. The aim of this stage is to obtain more robust empirical insights across different organisational contexts. The evaluation focuses on perceived usefulness, clarity of the framework and its contribution to prioritisation and decision-making. The data collection combines survey-based feedback, interviews and workshop-based discussions, where feasible.

The final phase represents an integration stage, where findings from all participating organisations are synthesised. This phase applies a mixed-method approach, combining descriptive statistical analysis with qualitative interpretation. The purpose is to identify common patterns, benefits and limitations of the framework and to refine the artefacts accordingly.

The validation process is inherently iterative. Feedback from each phase is used to adjust both the conceptual structure of the AIROF framework and the functionality of the AIROF Scorecard. This approach reflects the adaptive nature of the research and supports the development of a practically applicable governance framework for SME environments (Hevner et al., 2004; Peffers et al., 2007).

4. Discussion

Compared with existing governance frameworks, AIROF is not intended to replace legal compliance or formal management systems. Rather, it is designed to partially reduce the operational gap between high-level governance principles and day-to-day SME decision-making. The EU AI Act explains what is categorised and regulated; NIST explains how risks can be governed across lifecycle functions; ISO/IEC 42001 explains how to institutionalise AI management; AIROF is intended to support consulting SMEs in converting these expectations into a lightweight scorecard for routine use across the organisation. Table 1 presents a comparison of selected AI governance frameworks with AIROF.

Table 1. Comparison of AIROF with selected AI governance and risk frameworks

Framework	Main purpose	Main strength	Weakness for SME practice	What AIROF adds
EU AI Act	Legal framework based on risk levels	Binding rules and clear risk categories	Not an operational managerial tool for day-to-day decision-making	Translation of regulation into internal managerial decisions
NIST AI RMF	Voluntary AI risk management	Lifecycle logic of GOVERN–MAP–MEASURE–MANAGE	Can be too abstract for smaller firms	Simple evidence capture, scoring and ownership assignment
ISO/IEC 42001	AI management system	Institutional governance and continual improvement	Higher implementation burden	Lightweight entry tool before a full AIMS
AIROF	Practical management of AI risks and opportunities in SMEs	Role-based scorecard, adaptability, inclusion of opportunities	Still in the design and validation phase	Operationalisation of AI governance in daily practice

Source: author based on European Union (2024), NIST (2023), ISO (2023a, 2023b)

The academic contribution lies in linking fourth-order cybernetics with AI governance and practical artefact design. The managerial contribution lies in providing a structured yet low-threshold instrument for organisations that need actionable AI oversight but do not have the capacity to deploy a full management system immediately. The paper therefore seeks to contribute to both theory and practice, while acknowledging that the empirical validation is ongoing.

5. Conclusion

This paper has presented AIROF as a work-in-progress conceptual framework designed to support the adaptive management of risks and opportunities associated with the use of artificial intelligence in consulting SMEs. The framework is grounded in fourth-order cybernetics, which emphasises feedback, reflexivity and continuous adaptation in complex socio-technical systems.

By integrating selected principles of AI governance and risk management into a unified conceptual structure, AIROF is intended to partially address the operational gap between high-level frameworks and everyday managerial practice. At the present stage, this gap cannot be claimed to be fully closed, because empirical validation has not yet been completed. The contribution of this paper is therefore prospective: it presents the theoretical and methodological foundation of the framework and prepares the ground for its subsequent design, implementation and evaluation in consulting SMEs. While current approaches define what should be governed, they often lack mechanisms for how governance can be operationalised in everyday practice, particularly in resource-constrained SME environments.

The proposed framework contributes by conceptualising governance as a dynamic and evolving process rather than a static set of rules. This perspective reflects the increasing complexity and unpredictability of AI systems, where continuous feedback and iterative adjustment become essential for effective management.

At the current stage, the research is focused on finalising the AIROF framework and developing its operational form through the AIROF Scorecard. The next phase will involve deployment and

empirical validation in consulting SMEs, aiming to assess its usability, clarity and impact on decision-making processes.

From a theoretical perspective, the research contributes to the application of cybernetic principles in the field of AI governance. From a practical perspective, it aims to provide a lightweight and adaptable framework that supports organisations in navigating the evolving landscape of AI-related risks and opportunities.

Acknowledgement

Supported by grant No. IP400040 – Support for Doctoral Research at Prague University of Economics and Business.

References

Ashby, W. R. (1956). An introduction to cybernetics. Chapman & Hall.

Božičnik, S., & Mulej, M. (2011). A new–fourth order cybernetics and sustainable future. Kybernetes, 40(5/6), 670–684.

Capra, F., & Luisi, P. L. (2014). The systems view of life. Cambridge University Press.

Committee of Sponsoring Organizations of the Treadway Commission. (2017). Enterprise risk management: Integrating with strategy and performance. COSO.

Davenport, T. H., & Mittal, N. (2023). All-in on AI. Harvard Business Review Press.

European Union. (2024). Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). EUR-Lex. https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. MIS Quarterly, 28(1), 75–105.

International Organization for Standardization. (2018). ISO 31000: Risk management – Guidelines.

International Organization for Standardization. (2023a). ISO/IEC 23894:2023 Information technology—Artificial intelligence—Guidance on risk management. https://www.iso.org/standard/77304.html

International Organization for Standardization. (2023b). ISO/IEC 42001:2023 Information technology—Artificial intelligence—Management system. https://www.iso.org/standard/42001

Medvedeva, T. A., & Umpleby, S. A. (2024). Cybernetics of social systems. Systems Research and Behavioral Science.

National Institute of Standards and Technology. (2023). Artificial intelligence risk management framework (AI RMF 1.0) (NIST AI 100-1). U.S. Department of Commerce. https://nvlpubs.nist.gov/nistpubs/ai/nist.ai.100-1.pdf

Organisation for Economic Co-operation and Development. (2024). OECD AI principles. https://oecd.ai/en/ai-principles

Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). Design science research methodology. Journal of Management Information Systems, 24(3), 45–77.

Russell, S., & Norvig, P. (2021). Artificial intelligence: A modern approach (4th ed.). Pearson.

United Nations Educational, Scientific and Cultural Organization. (2021). Recommendation on the Ethics of Artificial Intelligence. https://unesdoc.unesco.org/ark:/48223/pf0000381137

von Foerster, H. (1981). Observing systems. Intersystems.

Wiener, N. (1948). Cybernetics. MIT Press.

ARTIFICIAL INTELLIGENCE AND TECHNO-STRESS AT WORKPLACES: A SCOPING REVIEW

Kamila Němečková, Petr Řehoř, Nikola Soukupová, Lukáš Klarner

University of South Bohemia in Ceske Budejovice
Faculty of Economics
nemeck07@ef.jcu.cz, rehor@ef.jcu.cz, nsoukupova@ef.jcu.cz, klarnl00@ef.jcu.cz

DOI: 10.35011/IDIMT-2026-409

Keywords

Artificial Intelligence; techno-stress; workplace; scoping review.

Abstract

With the increasing integration of artificial intelligence (AI) in workplaces, attention to techno-stress stemming from employees' use of AI has grown significantly. Techno-stress is a kind of stress that arises from exposure to digital technologies, such as mobile phones, computers, or AI. This exploratory scoping review aims to map and analyze the published English-language literature from 2015 to 2026 on technostress stemming specifically from AI use in workplaces, and to identify relevant research gaps. Relevant studies were retrieved from the Web of Science and Scopus databases. A total of 12 papers were included in the final sample, all of which were published between 2024 and 2026. The selected papers indicate the dual effect of AI techno-stress: both a challenge and a hindrance. The research also emphasizes mediators, such as job anxiety and work engagement, and moderators, such as technical self-efficacy and organizational support, and their positive and negative effects on AI technostress. These findings show the complex and context-dependent nature of AI techno-stress and highlight the factors that mediate and/or moderate techno-stress in workplaces.

1. Introduction

Nowadays, artificial intelligence (AI) is increasingly adopted by many companies. AI in companies faces new, unfamiliar problems, brings alteration of work content, and could possibly leave a negative effect on employees (Brynjolfsson et al., 2025; Zhou et al, 2024). Generative artificial intelligence (GenAI) based on large-language machine-learning models (LLMs) is gaining a lot of attention recently. It started to appear only in 2022 after OpenAI introduced ChatGPT-3.5 and gained popularity among employees of various industries. It also represents models such as Google´s Gemini, or Claude (Hoffmann et al, 2025; Al Naqbi et al., 2024). GenAI functions in cooperation with people and rarely operates without human supervision. It creates or modifies images, videos, or text, and it can also do problem-solving or decision-making, which creates a space for AI/human collaboration (Schecter & Richardson, 2025; Fui-Hoon Nah et al., 2023). AI in general creates challenges and opportunities in many fields, such as healthcare, business, and education, where it can be efficiently used (Fui-Hoon Nah, 2023). In companies, the collaboration of AI/employee or

AI/company can bring many positive effects, such as boosting employee productivity or bringing a competitive advantage. AI also has its negative side, for example, violation of personal privacy, misinformation, or distortion of social trust. For companies and their employees, the negative effects represent job insecurity or low willingness to work with AI tools (Sang-Hyeak et al., 2025; Hai et al., 2025). With the usage of AI comes the side effects on employees´ mental health and stress levels.

The type of stress a person is experiencing when using digitalization tools or AI tools is signified as techno-stress. Negative effects of AI usage could have a significant influence on the rise of employee techno-stress (Högemann et al., 2025). It stems from interactions between organizational design, technological complexity, and individuals' cognition and is a critical factor in human/technology interaction. When it comes to working and collaborating with AI in general, the context of techno-stress gains new and greater complexity (Jin et al., 2026). Employees who are experiencing techno-stress stemming from AI use often experience pressure to increase workload or work pace (Zhou et al., 2024). Therefore, it is necessary to investigate and research the influence of AI on the stress levels of employees and identify which techno-stressors are relevant in the context of AI (Issa et al., 2024). The current literature is focused mainly on digitalization in general or on digital transformation in the context of techno-stress and techno-stressors; it is also necessary to understand the role that AI is playing in this ongoing topic, which many companies are starting to deal with only recently.

The main aim of this paper is to analyse the current literature published in 2015-2026 in chosen databases regarding technostress and Artificial Intelligence (AI) used in workplaces. The analysis itself is conducted for a better understanding of the ongoing research on AI and technostress, which stems from AI usage by employees in workplaces, and for the identification of research gaps. The findings are to be used in future research, which will focus on technostress in workplaces in general, in which AI could play an important role. The successful fulfilment of the main aim is set following the research question (RQ):

RQ: What are the trends in literature from 2015 to 2026 dealing with the use of AI by employees and the technostress (or techno-stressors) resulting from its use?

2. Methods

The methodology of this explorative scoping review was conducted according to several authors (Peters et al., 2020; Munn et al., 2018) and PRISMA-ScR guidelines. PRISMA-ScR is designed to help understand the key concepts and terminology related to a selected topic. It is also intended to assess the scope and characteristics of current literature focused on a specific topic (Tricco et al., 2018). First, the RQ was determined, as can be seen above. Subsequently, the criteria for the first search were defined according to the recommendation of the Joanna Brigs Institute for the inclusion and exclusion criteria identification. The PCC framework will be used, which includes three factors – population, concept, and context (Peters et al., 2020). The PCC framework set for this paper is shown in Table 1.

Table 1. PCC framework

Criteria	Characteristics
Population	Employees with all employee statuses from all industries
Concept	AI and techno-stress related to AI usage
Context	All countries, only articles or reviews published between 2015 and 2026 in English language

Source: authors

As can be seen in Table 1, the papers which were included for the search are only articles or reviews published between 2015 and 2026 in the English language. Restrictions on countries or industries were not applied. The population includes employees with all employee statuses working in all industries, and the concept includes AI and technostress, which stems from AI usage. The databases used for the paper search included Web of Science (WoS) and Scopus. The next steps in the whole process included the identification of relevant papers and paper selection. The paper search included a searching algorithm based on the set PCC framework, which is shown in Table 2.

Table 2. Search algorithms according to PCC

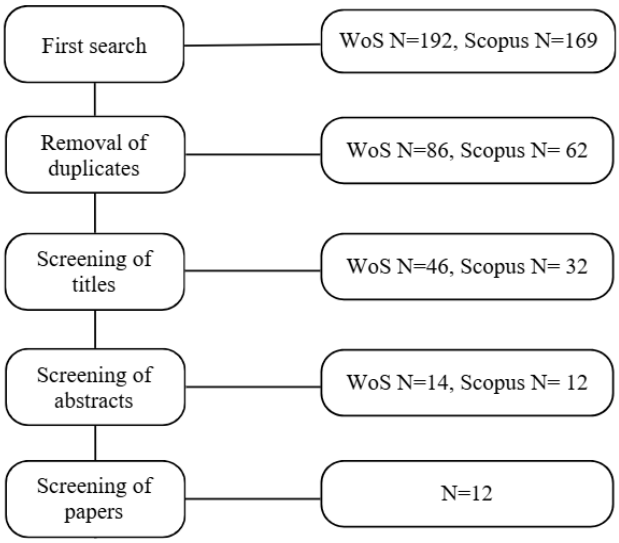
Criteria	Search algorithms and keywords
Population	(“employee*” OR “workplace” OR “organization*” OR “work environment”)
Concept	(“technostress” OR “techno-stress” OR “technology-induced stress” or “digital stress”) AND (“artificial intelligence” OR “AI” OR “algorithmic management” or “digital transformation” or “intelligent systems”)
Context	Only filters applied – year of publication and document type, language

Source: authors

Table 2 describes the first search algorithm and keywords. Keywords were identified based on the mentioned PCC framework criteria. Under the “population” criterion, keywords related to organization and employees were identified. Under the “concept” criterion, keywords associated with techno-stress and, directly or indirectly, with AI were identified. No specific keywords were specified for the “context” criterion. There are keywords that are not directly related to AI in companies, such as “algorithmic management” or “digital transformation”. However, these keywords were included because there are discussions about AI implementation that occur within these specific topics.

The first search was followed by a selection of papers, which is shown in Figure 1. First, basic filters were applied (language, publication year, and type of paper). Subsequently, on the 26th of February 2026, the article information was downloaded, and duplicates were removed. This was followed by a scan of the article titles, abstracts, and full texts. During this phase, it was important to identify articles related to AI in various work environments and its impact on employee techno-stress. Articles that focused solely on general digitalization without mentioning AI were excluded. The total number of papers used for this scoping review is 12.

Figure 1. Process of paper selection



Source: authors

The next step involved data extraction and data synthesis. Following this step, the results were analysed, and the data were interpreted and discussed.

3. Results

For this explorative scoping review, 12 papers – articles or reviews – were chosen according to the PCC framework of paper selection. This chapter will present the basic characteristics and main themes of the chosen papers.

3.1. Basic characteristics of chosen papers

Basic characteristics of selected papers will be shown first in Table 3 below. The Table itself contains information about the author, the year of publication, the country of research conduction, research topic, design or methods which were mainly used to obtain the data, and the sample used in the paper.

Table 3. Basic characteristics of selected papers

Author, Year of publication	Country	Topic/focus	Design/methods	Sample
Sharif et al., 2025	China	AI influence on employee work and career development	3-wave online survey w/ 2-week interval between waves	277 employees of chain restaurants
Zhang et al., 2025	China	Technostress and employee innovative behaviour in the era of AI	2-wave questionnaire w/ 1-month interval between waves	381 employees of manufacturing industry
Nayak et al., 2025	India	AI interventions influence on HR in hotels	2-wave interviews	55 HR practitioners working at hotels
Issa et al., 2024	USA	Effects of techno-stressors within the context of AI	E-survey questionnaire	224 professionals from healthcare
Zhou et al., 2024	China	Knowledge and skill development in digital transformation and AI	1-wave survey	224 employees of car-testing services
Chang et al., 2024	China	Challenge/hindrance techno-stressors and intention to adopt AI	3-wave survey w/ 1-month interval between waves	301 employees from various sectors
Liu et al., 2025	China	Mechanism of AI job crafting and medical staff presenteeism	1-wave online survey	225 participants - medical staff
Lopes et al., 2026	Portugal	Psychosocial predictors, technostress and perceived usefulness of AI	1-wave online survey	422 respondents
Högemann et al., 2025	Germany	Perception of technostress in the context of GenAI by young professionals	Semi-structured interview	15 young professionals

Author, Year of publication	Country	Topic/focus	Design/methods	Sample
Jia et al., 2025	China	Perceived values, responses and attitudes towards GenAI	Semi-structured interview	18 creative professionals
Feng et al., 2025	China	Sustainable AI adoption through mediating roles of technostress and innovation resistance	1-wave questionnaire	426 employees from various industries
Jin et al., 2026	China	AI pressures and techno-stressors and fostering/hindering of career development	2-wave field survey w/ 3-week interval between waves	326 employees from a state-owned research institute

Source: authors

As can be seen in Table 3, all the chosen papers were published between 2024 and 2026, which shows the popularity of the topic in recent years. Most of the papers used a multi-wave questionnaire/survey or interviews, and the country where the research was conducted was mostly China. It is also important to mention that all the selected papers are articles only, no reviews. The industries in which the research was conducted are various and include mainly hospitality, manufacturing, and healthcare. What also needs to be stressed is that most of the selected papers focus on AI in general; there are a few papers that particularly stress the importance of GenAI in workplaces.

3.2. Main findings

Through the papers, multiple main themes could be seen. Firstly, the dual nature of technostress when employees use AI – the positive and negative effects of technostress. Secondly, the role of mediators and moderators, and thirdly, AI-related stressors and their impact on work and the behaviour of employees. There could also be additional patterns seen across the studies. Mainly, the recognition of the complex nature of AI-related technostress with both positive and negative outcomes, which are reported simultaneously. The research also distinguishes between challenge and hindrance techno-stressors and claims that each type leads to different outcomes (Zhang et al., 2025; Chang et al., 2024). Lastly, several papers stress the importance of organizational context and support in shaping employees´ behaviour towards AI usage. Several papers also use the categorization of techno-stressors into *techno-overload, techno-invasion, techno-complexity, techno-uncertainty and techno-insecurity* and consecutively list the activities and factors which cause techno-stress when using AI (Zhang et al., 2025; Nayak et al., 2025; Issa et al., 2024; Högemann et al., 2025) Moreover, one paper suggests the extension of these five techno-stressor categories by adding a sixth category which is called *techno-unpredictability* and seems to be more significant in the context of AI than techno-uncertainty because of disruptive technologies (Issa et al., 2024). The findings suggest that techno-stress has a dual effect on employee outcomes, acting both as a challenge and a hindrance (Nayak et al., 2025). Moreover, the findings also suggest that techno-complexity and techno-insecurity are related to techno-eustress, which is the positive and motivating type of stress; techno-invasion and techno-unpredictability are positively related to techno-distress – the negative type of stress which can cause burnout (Issa et al., 2024). Positive effects of techno-stress are mainly associated with challenge-related techno-stressors and can cause increased work engagement, learning and innovative behaviour. It was also linked to improved job performance and career growth through mechanisms like learning and proactive employee behaviour. On the other hand, negative techno-stress stems from hindrance-related techno-stressors, which include mainly job anxiety, job insecurity, presenteeism and reduced well-being (Zhang et al., 2025; Sharif et al., 2025). It was also associated with lower

intention of AI adoption and negative emotional responses towards AI or anxiety in general (Jia et al., 2025). This stresses the different contexts and nature of technostress when using AI and its effects. Several studies also identified mediating or moderating factors which can shape the effects of technostress. Job anxiety or job insecurity mediated technostress in a negative way and had a negative impact on well-being or career outcomes (Sharif et al., 2025). On the other hand, work engagement or learning could have a positive effect and could transform AI-related stress into better performance and supportive behaviour (Zhou et al., 2024). Moderators which were identified include organizational support, technical self-efficacy and AI trust, organizational culture and digital responsibility or leadership behaviours (Zhou et al., 2024; Avtalion et al., 2025). These moderators could reduce techno-stress, increase AI adoption or moderate the relationship between AI usage and employees´ outcomes and behaviour.

4. Discussion

This scoping review aimed to map the current literature on techno-stress stemming from AI use across various workplaces. Several main findings should be highlighted. First, there could be seen multiple main themes across the papers – the dual nature of techno-stress stemming from AI use (with both positive and negative outcomes reported simultaneously), the important role of moderators and mediators when it comes to the perception of AI techno-stress and AI-related techno-stressors which have an impact on employees´ work and behaviour (Sharif et al, 2025, Zhang et al., 2025). It was also mentioned that AI techno-stress is a specific category of stress and therefore, could introduce a new dimension to techno-stress perception – techno-unpredictability (Issa et al., 2024). What also needs to be stressed is the importance of organizational context and the AI implementation strategy in managing techno-stress stemming from AI use – organizational support, leadership, self-efficacy, and resilience could play an important role in this. There is a large variety of techno-stress interventions that could be suggested. Sharif et al. (2025) suggested that companies that implement AI should foster an environment that encourages employee development. This could also include employee adaptation to develop basic AI skills and knowledge. Regarding AI unpredictability, Issa et al. (2024) suggest that uncertainty management grounded in explanations of AI predictions could lead to a better understanding of AI techno-stressors. Högemann et al. (2025) suggest implementing AI based on employees' skills and knowledge to empower their development. This aligns with the suggestion of Chang et al. (2024), who stated that developing employees' AI skills can reduce negative emotions towards AI and rewarding employees who show interest in developing their AI skills should empower positive emotions towards AI. For individual stress measurement, companies could use, for example, wearable devices that measure physical signs of stress (Can et al., 2019). Despite the growing popularity of this topic, there may still be research gaps. The literature is focused mainly on specific industries – healthcare or hospitality. There are still industries that lack research on this topic. Subsequently, the research has been mainly centred in China over the last 10 years, and there is a lack of literature in other regions, such as European countries. Future research, therefore, could focus on the context of European companies and their current approaches to techno-stress stemming from AI. The next area of focus could include interventions to prevent or limit the negative effects of AI techno-stress. This could open the discussion on a suitable organizational environment that protects employees from the negative effects of AI techno-stress and supports their development in AI use. This scoping review has several limitations. The research focused solely on English-language written literature (mainly articles and reviews) published between 2015 and 2026. The current literature is focused on specific sectors and countries, primarily hospitality and healthcare in China. Subsequently, the scoping review aimed to map the existing literature rather than provide an in-depth critical appraisal, which may limit the depth of the analysis, and the number of papers was limited to 12.

5. Conclusion

To answer the research question, several trends were found in the field of technostress stemming from AI use in workplaces. These trends included the dual nature of techno-stress, the roles of mediators and moderators in techno-stress in the context of AI, and the impact of AI-related techno-stressors on work and employee behaviour. Also, several other patterns emerged in the research. These include the complex nature of techno-stress, the categorization of challenge-related and hindrance-related techno-stressors, techno-overload, techno-invasion, techno-complexity, techno-uncertainty, techno-insecurity, and techno-unpredictability, and, lastly, the organizational context and its effects on techno-stress in the context of AI. This scoping review aimed to analyse the current literature on technostress and AI in workplaces and to identify research gaps that can be marked as fulfilled.

Acknowledgement

This paper was supported by Grant Agency of University of South Bohemia in České Budějovice GA JU 022/2026/S.

References

Al Naqbi, H., Bahroun, Z., & Ahmed, V. (2024). Enhancing Work Productivity through Generative Artificial Intelligence: A Comprehensive Literature Review. Sustainability, 16(3), 1166. https://doi.org/10.3390/su16031166

Brynjolfsson, E., Li, D., & Raymond, L. (2025). Generative AI at Work. The Quarterly Journal of Economics, 140(2), 889-942. https://doi.org/10.1093/qje/qjae044

Can, Y. S., Chalabianloo, N., Ekiz, D., & Ersoy, C. (2019). Continuous Stress Detection Using Wearable Sensors in Real Life: Algorithmic Programming Contest Case Study. Sensors, 19(8), 1849. https://doi.org/10.3390/s19081849Ř

Chang, P. C., Zhang, W., Cai, Q., & Guo, H. (2024). Does AI-Driven Technostress Promote or Hinder Employees' Artificial Intelligence Adoption Intention? A Moderated Mediation Model of Affective Reactions and Technical Self-Efficacy. Psychology Research and Behavior Management, 17, 413-427. https://doi.org/10.2147/prbm.s441444

Feng, Y., Feng, Y., & Liu, Z. (2025). The Influence of Perceived Organizational Support on Sustainable AI Adoption in Digital Transformation: An Integrated SEM–ANN–NCA Model. Sustainability, 17(24), 11373. https://doi.org/10.3390/su172411373

Fui-Hoon Nah, F., Zheng, R., Cai, J., Siau, K., & Chen, L. (2023). Generative AI and ChatGPT: Applications, challenges, and AI-human collaboration. Journal of Information Technology Case and Application Research, 25(3), 277-304. https://doi.org/10.1080/15228053.2023.2233814

Hai, S., Long, T., Honora, A., Japutra, A., & Guo, T. (2025). The dark side of employee-generative AI collaboration in the workplace: An investigation on work alienation and employee expediency. International Journal of Information Management, 83, 102905. https://doi.org/10.1016/j.ijinfomgt.2025.102905

Hoffmann, M., Boysel, S., Nagle, F., Peng, S., & Xu, K. (2025). Generative AI and the nature of work (Harvard Business School Strategy Unit Working Paper No. 25-021; Harvard Business School Working Paper No. 25-021; CESifo Working Paper No. 11479). SSRN. https://doi.org/10.2139/ssrn.5007084

Högemann, M., Hein, L., Britsche, J. -O., & Thomas, O. (2025). Technostress and generative AI in the workplace: a qualitative analysis of young professionals. Frontiers in Artificial Intelligence, 8, 1728881. https://doi.org/10.3389/frai.2025.1728881

Issa, H., Jaber, J., & Lakkis, H. (2024). Navigating AI unpredictability: Exploring technostress in AI-powered healthcare systems. *Technological* Forecasting and Social Change, 202, 123311. https://doi.org/10.1016/j.techfore.2024.123311

Jia, M., Zhao, Y. C., & Zhang, X. (2025). Will potential threats persist? Attitudes and behavioral responses of creative professionals toward GenAI. Information Research an international electronic journal, 30, 1073-1083. https://doi.org/10.47989/ir30iconf47098

Jin, T., Yang, X., & Zhang, L. (2026). Understanding AI Technostress and Employee Career Growth from a Socio-Technical Systems Perspective: A Dual-Path Model. Systems, 14(1), 58. https://doi.org/10.3390/systems14010058

Liu, R., Deng, W., Wang, T., & Yang, T. (2025). Enabler or inhibitor? The effect of AI crafting on presenteeism among medical staff. European Journal of Work and Organizational Psychology, 35(2) 186-200. https://doi.org/10.1080/1359432x.2025.2492632

Lopes, J. M., Gomes, S., Nogueira, E., Ferreira, J. J., & Dabić, M. (2026). Exploring psychosocial predictors of AI use intention: the moderating role of organisational culture in diverse workplaces. Behaviour & Information Technology, 1-19. https://doi.org/10.1080/0144929X.2026.2626743

Munn, Z., Peters, M. D. J., Stern, C., Tufanaru, C., McArthur, A., & Aromataris, E. (2018). Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. BMC Medical Research Methodology, 18(1), 143. https://doi.org/10.1186/s12874-018-0611-x

Nayak, S., Budhwar, P., & Malik, A. (2025). Unveiling the hidden costs of AI in hospitality industry: The adverse effect of algorithmic human resources management (AHRM) on stress, wellbeing, and commitment. International Journal of Hospitality Management, 129, 104206. https://doi.org/10.1016/j.ijhm.2025.104206

Peters, M. D. J., Marnie, C., Tricco, A. C., Pollock, D., Munn, Z., Alexander, L., McInerney, P., Godfrey, C. M., & Khalil, H. (2020). Updated methodological guidance for the conduct of scoping reviews. JBI Evidence Synthesis, 18(10), 2119-2126. https://doi.org/10.11124/jbies-20-00167

Schecter, A., & Richardson, B. (2025). How the Role of Generative AI Shapes Perceptions of Value in Human-AI Collaborative Work. Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems, Article 530, 1-15. https://doi.org/10.1145/3706598.3713946

Sharif, M. N., Zhang, L., Asif, M., Alshdaifat, S. M., & Hanaysha, J. R. (2025). Artificial intelligence and employee outcomes: Investigating the role of job insecurity and technostress in the hospitality industry. Acta Psychologica, 253, 104733. https://doi.org/10.1016/j.actpsy.2025.104733

Tricco, A. C., Lillie, E., Zarin, W., O'Brien, K. K., Colquhoun, H., Levac, D., Moher, D., Peters, M. D. J., Horsley, T., Weeks, L., Hempel, S., Akl, E. A., Chang, C., McGowan, J., Stewart, L., Hartling, L., Aldcroft, A., Wilson, M. G., Garritty, C., et al. (2018). PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. Annals of Internal Medicine, 169(7), 467-473. https://doi.org/10.7326/m18-0850

Zhang, S., Guo, P., Yuan, Y., & Ji, Y. (2025). Anxiety or engaged? Research on the impact of technostress on employees' innovative behavior in the era of artificial intelligence. Acta Psychologica, 259, 105442. https://doi.org/10.1016/j.actpsy.2025.105442

Zhou, Q., Chen, K., & Cheng, S. (2024). Bringing employee learning to AI stress research: A moderated mediation model. Technological Forecasting and Social Change, 209, 123773. https://doi.org/10.1016/j.techfore.2024.123773

SELECTED PERSPECTIVES ON REPUTATIONAL RISK IN THE CONTEXT OF CYBERSECURITY

Romana Sobotová

Bratislava University of Economics and Business, Slovakia
Faculty of Business management
romana.sobotova@euba.sk

František Pollák

Institute of Technology and Business in Ceske Budejovice
Faculty of Corporate Strategy
frantisek.pollak@vste.cz

DOI: 10.35011/IDIMT-2026-417

Keywords

Cyber incidents; reputation management; cybersecurity; digital trust; reputational risk; crisis communication.

Abstract

This paper reviews recent literature on corporate reputation, digital trust, crisis communication, and cyber incidents, and complements it with a comparative analysis of Eurostat data for EU-27 countries. The paper focuses on cybersecurity-related incidents leading to data destruction or corruption. The results show that reputational vulnerability in the digital environment is shaped not only by the occurrence of cyber incidents, but also by the extent to which such incidents can undermine stakeholder trust and perceptions of organizational reliability. In the case of Slovakia, the findings indicate relatively low reported incident rates compared with the EU average. The paper concludes that in the context of cybersecurity, reputational risk depends not only on the occurrence of incidents, but also on the ability of organizations to respond to them effectively.

1. Introduction

Digitalization has changed the conditions under which organizations build and lose reputational capital. As stakeholders interact with organizations through social media platforms, algorithmic information flows, and real-time digital channels, reputation management has become materially more complex and severe. Traditional approaches based on managed messaging, stable communication channels, and a strong managerial grip on the corporate narrative have proven to be insufficient in environments where information spreads instantly, narratives are challenged by multiple actors, and adverse events reach global reach within hours (Krstić, 2025; Badham & Luoma-aho, 2023; Bartekova & Janikovicova, 2026). Contemporary literature increasingly places reputation management not only as a function of public relations or communication, but as a strategic

management discipline with direct impacts on organizational performance, stakeholder trust, resilience, and long-term value creation (Baumgartner et al., 2020; Krstić, 2025). This review synthesizes scientific knowledge from strategic management, stakeholder theory, digital transformation research, and business communication to explore how digital acceleration is transforming reputation creation, stakeholder dynamics, and organizational risk. The following sections systematically address these transformations and conclude with implications for managerial practice and directions for future research.

2. Related works

Corporate reputation is a multidimensional, stakeholder-constructed assessment of an organization that is shaped over the long term through observed behaviour, experience, and mediated communication (Krstić, 2025; Chaudhri et al., 2021). From the point of view of resource theory, it is a valuable and difficult to imitate asset generating a competitive advantage and reducing stakeholder uncertainty (Baumgartner et al., 2020). Reputation management is not just a communication function, but a strategic management discipline encompassing leadership, governance, crisis response, and cybersecurity. Vassallo et al. (2024) demonstrate in the Journal of Management that reputation is subject to cumulative damage from recurring adverse events, requiring continuous, not episodic, recovery mechanisms. This perspective is also consistent with earlier and regionally relevant research. Pollák et al, (2021) explicitly treat corporate reputation as a key success factor and as an intangible asset whose protection requires continuous monitoring, particularly in the online environment; at the same time, their analysis of family-owned businesses shows that meaningful differences may exist between the online reputation of parent companies and that of their brands. A related contribution is offered by Dorčák et al. (2014), who addressed the possibilities of improving the online reputation of public institutions and approached online reputation as a phenomenon that can be actively analysed, managed, and improved in the digital environment rather than passively observed. This line of thinking is complemented by Markovič et al. (2022), who, in the context of the Slovak e-commerce market, showed that online interactions and consumer behaviour can change rapidly under external shocks and subsequently stabilise into new behavioural patterns. For the present study, this is important because it reinforces the assumption that digital signals, stakeholder responses, and reputational exposure in online environments should be monitored continuously rather than assessed only retrospectively. Accelerated digitization has fundamentally changed the conditions under which reputation is formed. Social media and online platforms have become the primary arenas where stakeholders observe, evaluate, and publicly interpret organizations, effectively ending organizations' monopoly on self-image creation (Badham & Luoma-aho, 2023). User-generated content (UGC) and algorithmic visibility reinforce the reach of negative information that spreads quickly and persists over the long term (Rust et al., 2021; Chung et al., 2019). Cyber incidents are a particularly serious category: Confente et al. (2019) have shown that data leaks generate UGCs framing such events as a deficit in organizational integrity. Chan et al. (2025) have shown how ideologically charged digital crises can escalate through algorithmic amplification into uncontrollable reputational threats. Stakeholder trust is a key mediator between a cyber incident and its reputational consequences. Tong and Chan (2022) emphasize that digital trust depends on the quality of dialogic communication and the perceived consistency of organizational behaviour. Başaran (2024) found that proactive crisis communication strategies are significantly more effective in protecting reputation than reactive responses. Nuortimo et al. (2024) as well as Janikovičová and Barteková (2026) confirms that digital environments make crisis communication more critical, as the asymmetry between the speed of crisis escalation and the capacity of organizational response has widened significantly. Eckhardt (2020) points out that in the reputation economy, peer-generated institutional

credentials are increasingly being replaced, making the cumulative quality of stakeholder experience the primary source of reputational advantage. Effective reputation management in the digital age requires a shift from a reactive to a proactive approach. Rust et al. (2021) argue that reputation is a continuously changing signal, which requires continuous analytical monitoring. In a systematic review of 104 studies, Molavi and Zhang (2025) affirm that AI tools enable real-time monitoring of reputation signals, but also highlight ongoing challenges in scalability, algorithm ethics, and reputation measurement at the organizational level.

3. Methodology

The empirical part of the paper is based on a secondary comparative quantitative analysis. It aims to assess the extent to which businesses in the EU-27 are exposed to cyber incidents with a potential reputational impact. Since publicly available databases do not contain a direct indicator of corporate reputation, the analysis does not measure reputation directly but focuses on the occurrence of ICT security incidents leading to the destruction or corruption of data, which is used in this paper as a proxy indicator of reputational risk. In other words, the selected indicator does not capture reputational damage itself, but one type of cyber event that may weaken stakeholder trust and perceptions of organizational reliability.

The unit of analysis is the EU-27 member states. The examined data relate to enterprises with 10 or more employees and are based on the Eurostat – Information Society Statistics (ESTAT) database. Undertakings from most sectors are included in the analysis, excluding the agricultural, forestry, fisheries, mining and financial sectors, in line with Eurostat's methodology. This exclusion should be treated as an important limitation of the study, since financial institutions, and especially banks, are highly relevant from the perspective of both cybersecurity and reputational risk.

The monitored indicator captures the share of enterprises that have experienced an ICT security incident leading to the destruction or corruption of data (Eurostat, isoc_cisce_icn2). The choice of this indicator was guided mainly by data availability and cross-country comparability within the Eurostat dataset. Although incidents involving the disclosure of confidential data may have a more direct and visible reputational impact, the selected indicator provides a consistent basis for comparison across EU countries and still reflects a serious breach of data integrity with potential reputational consequences. Data are available for 2022 and 2024 and allow for cross-country comparisons at EU-27 level, including a reference average. The Eurostat methodology used in this paper refers to enterprises with 10 or more employees and provides harmonised cross-country data for selected ICT security indicators, which makes it suitable for comparative analysis at the EU level.

In terms of time, this is a comparison of two measuring points, 2022 and 2024, which makes it possible to capture the basic direction of development and current inter-country differences. Descriptive and comparative statistical procedures are used: differences in percentage points, the ranking of countries within the EU-27 and the development of the indicator in Slovakia against the European average.

The advantage of the chosen methodology is that it works with harmonized and internationally comparable Eurostat data, which allow Slovakia to be placed in a wider European context. It should be emphasized that the analysed indicator measures reported incidents, which may underestimate the actual incidence, since many incidents remain unreported and create a dark figure of cyber events that is not captured in official statistics. This is important in the context of reputational risk, as organisations may avoid disclosure to limit possible damage to stakeholder trust and public credibility. At the same time, incident reporting is increasingly shaped by legal and regulatory

obligations, including data protection and cybersecurity rules such as GDPR and the broader reporting requirements associated with the NIS2 framework.

4. Results

Cyber incidents leading to the destruction or corruption of data represent an important source of reputational risk in the digital environment. When data integrity is compromised, the consequences extend beyond the technical sphere: such events may undermine stakeholder trust, weaken perceptions of organisational competence, and raise doubts about the organisation's ability to manage digital resources responsibly. In line with prior research, the reputational impact of cyber incidents often outlasts the incident itself, as negative information may continue to circulate online even after the technical problem has been resolved (Confente et al., 2019; Chan et al., 2025). At the EU-27 level, 5.2% of enterprises reported an ICT security incident leading to data destruction or corruption in 2022. By 2024, this value had declined slightly to 5.0%, suggesting a rather stable than substantially improving situation. Although the EU average appears moderate, country-level differences are considerable. This indicates that reputational vulnerability linked to cyber incidents is unevenly distributed across the Union and is likely shaped by differences in digital exposure, cybersecurity governance, reporting practices, and overall organisational preparedness. As shown in Figure 1, the highest levels of reported incidents in 2024 were concentrated mainly in Central and Eastern Europe. Poland and Czechia were among the countries with the highest shares of enterprises affected by incidents of this type, while Romania also remained above the EU average. By contrast, some Western and Northern European countries, such as Ireland, Luxembourg, and Estonia, recorded lower levels. The map does not suggest a perfectly uniform regional pattern, yet it clearly points to a stronger concentration of reported reputationally relevant cyber incidents in the eastern part of the EU. From the perspective of reputational risk, these countries can be considered more exposed to events capable of disrupting trust, operations, and the perceived reliability of organisations.

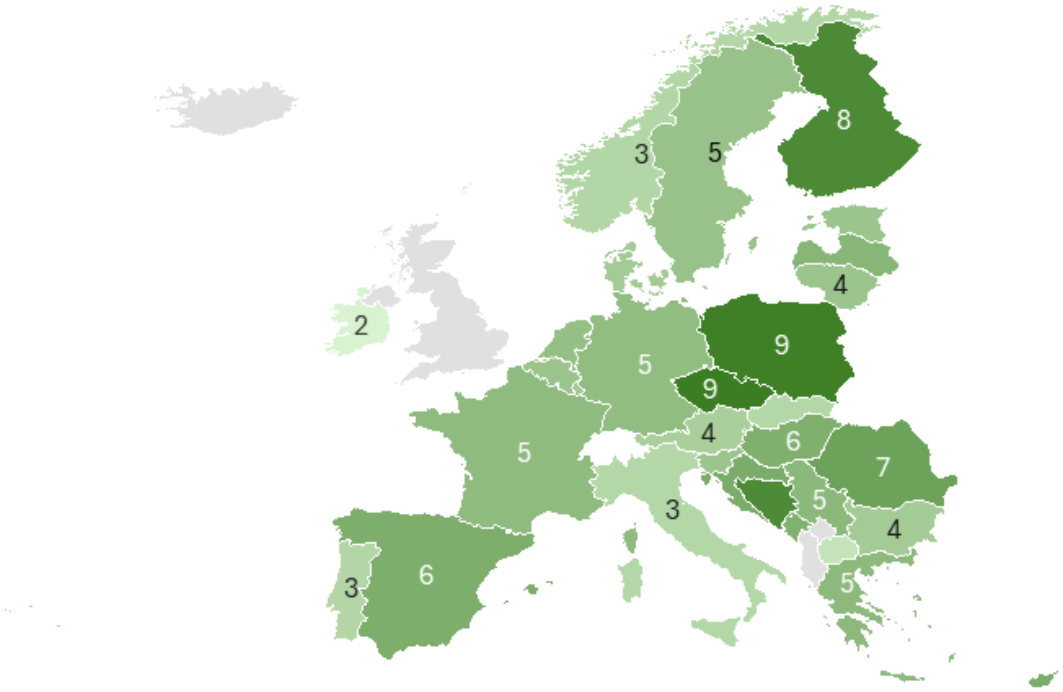
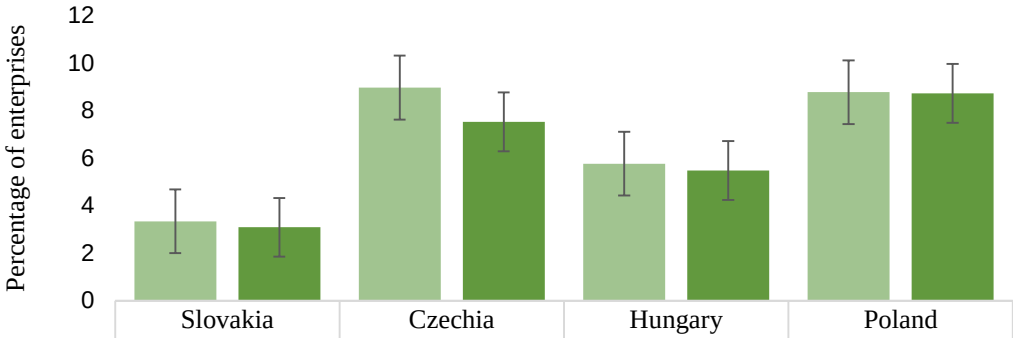


Figure 1. Enterprises that have experienced an ICT incident leading to data destruction or corruption – share (%) in the EU 2024
Source: Eurostat

Slovakia reported a lower rate of such incidents than the EU average in both observed years. The share of enterprises affected decreased slightly from 3.3% in 2022 to 3.1% in 2024, which places Slovakia below the EU-27 average and, at first sight, in a more favourable position. From the viewpoint of reputation management, this may indicate lower immediate exposure to incidents that could damage stakeholder trust or harm perceptions of organisational competence. However, this interpretation should remain cautious. A lower reported incidence does not necessarily imply lower actual exposure, since cyber incidents may be underreported for legal, strategic, or reputational reasons.



	Slovakia	Czechia	Hungary	Poland
2022	3	9	6	9
2024	3	8	5	9

Figure 2. Enterprises that have experienced an ICT incident leading to data destruction or corruption – share (%) in the V4 countries 2022 and 2024 Source: Eurostat

The V4 comparison presented in Figure 2 provides a more nuanced regional perspective. Slovakia recorded the lowest proportion of reported incidents among the V4 countries in both years. Czechia showed a decline between 2022 and 2024, from roughly 9% to 8%, while Hungary also improved slightly, moving from around 6% to 5%. Poland remained the most exposed country in the V4 group, with values close to 9% in both observed years. In comparative terms, Slovakia appears less affected by severe data-related ICT incidents than its regional peers. At the same time, the graph suggests that reputational exposure to cyber incidents remains a relevant issue across the whole V4 region, although with different intensity.

5. Discussion

The results of the analysis confirm that cybersecurity represents a relevant dimension of the reputational vulnerability of companies in the digital environment. A useful structured reference in this context is the ENISA Threat Landscape, which provides an overview of major cyber threats, threat actors, attack patterns, and mitigation approaches in the EU. In the context of this paper, such a framework helps interpret ICT incidents not only as isolated technical failures, but as part of a broader threat environment that may affect organisational credibility, continuity, and stakeholder trust. The rate of reported ICT incidents leading to data destruction or corruption stands at 5.0% (2024) in the EU-27, with significant differences between countries based on the level of digital exposure, security governance and incident records.

Slovakia scores relatively more favourably than the EU-27 average in this indicator, 3.1% (2024) is lower than the European average of 5.0%, while it has decreased slightly since 2022 (3.3%). From a reputation management perspective, this indicates lower immediate exposure to stakeholder trust incidents. However, this conclusion should be interpreted with a grain of salt: lower reported accidentality may not reflect actual risk, as many businesses do not report incidents for fear of

reputational and legal repercussions, which is itself reputational risk behaviour (Confente et al., 2019; Başaran, 2024).

6. Conclusion

Reputation management in the context of accelerated digitalization requires special attention to cybersecurity as a factor directly affecting stakeholder trust. Digital environments accelerate the dissemination of incident information, strengthen the impact of user-generated content, and create lasting digital records that maintain reputational damage long after the issue has been technically resolved (Chaudhri et al., 2021; Chan et al., 2025).

An analysis of Eurostat data indicates that Slovakia's share of reported ICT incidents leading to data destruction or corruption is below the European average (3.1% compared to 5.0% of the EU-27 in 2024). This result can be partly interpreted as a favourable signal, but the lower reported accidentality may also reflect a weaker incident record, not necessarily a lower real risk.

Vassallo et al. (2024) point out that reputational resilience is not created by a one-time response to a crisis, but by building prevention and recovery mechanisms over the long term. In the context of cybersecurity, this suggests that consistent security governance and prepared crisis communication may provide an important basis for maintaining stakeholder trust in the event of an incident. Tong and Chan (2022) emphasize that authentic and consistent communication is key to maintaining trust, which takes on particular importance when a cyber incident undermines the perceived integrity of an organization. In the present paper, however, these factors are treated as theoretical implications rather than as variables directly tested in the empirical analysis, which focuses on reported ICT incidents as a proxy indicator of reputational risk.

At the same time, the findings also support the argument of Confente et al. (2019), according to which cyber incidents have reputational consequences beyond the technical level of the problem itself, as they undermine the perception of credibility and competence of the organization. Although the lower reported incidence of incidents in Slovakia is favourable, in connection with deficits in security governance, it cannot be ruled out that some organizations remain vulnerable latently rather than visibly. This confirms that reputation management in the digital environment cannot be evaluated only by the occurrence of incidents, but also by the level of preparedness that allows incidents to be prevented and managed effectively.

For managerial practice, this implies that reputational resilience in the context of cybersecurity depends on the consistency of technological measures, internal governance, employee awareness and preparedness for crisis communication. Future research should focus on longitudinal monitoring of the relationship between the level of security governance and the real-world reputational consequences of incidents across different institutional contexts.

Acknowledgement

This article is one of partial outputs of the currently implemented research grant VEGA no. 1/0110/24 and project of Bratislava University of Economics and Business in Bratislava no I-26-103-00.

References

Badham, M., & Luoma-aho, V. (2023). Introduction to the handbook on digital corporate communication. In M. Badham & V. Luoma-aho (Eds.), Handbook on digital corporate communication. Edward Elgar Publishing.

Basaran, D. (2024). Exploring the impact of social media crisis management on customer trust and corporate reputation: The case of Starbucks Türkiye. Research Journal of Business and Management, 11(2), 94–111.

Bartekova, M., & Janikovicova, S. (2026). Green innovations and their barriers in small and medium-sized enterprises in Central Europe. Marketing and Management of Innovations, 17(1), 1–13.

Baumgartner, K. T., Ernst, C. A., & Fischer, T. (2020). How corporate reputation disclosures affect stakeholders' behavioural intentions: Mediating mechanisms of perceived organisational performance and corporate reputation. Journal of Business Ethics, 175(2), 361–389.

Chan, T. K. H., Lee, Z. W. Y., Pan, M., & Sun, K. (2025). Understanding the drivers and outcomes of ideologically charged social media firestorms: The sociotechnical and social learning perspectives. Journal of Management Information Systems, 42(3), 737–766.

Chaudhri, V., Oomen, T. A. P., Pridmore, J., & Joon, A. (2021). "CARE" in social media: Perceptions of reputation in the healthcare sector. Journal of Communication Management, 25(2), 125–141.

Chung, S., Chong, M., Chua, J. S., & Na, J.-C. (2019). Evolution of corporate reputation during an evolving controversy. Journal of Communication Management, 23(1), 52–71.

Confente, I., Siciliano, G. G., Gaudenzi, B., & Eickhoff, M. (2019). Effects of data breaches from user-generated content: A corporate reputation analysis. European Management Journal, 37(4), 492–504.

Dorcak, P., Pollák, F., & Szabo, S. (2014). Analysis of the possibilities of improving an online reputation of public institutions. In Proceedings of the 22nd Interdisciplinary Information Management Talks: IDIMT-2014 Networking Societies—Cooperation and Conflict, Poděbrady (pp. 275–282).

Eckhardt, G. M. (2020). Playing the trust game successfully in the reputation economy. NIM Marketing Intelligence Review, 12(2), 10–17.

European Union Agency for Cybersecurity (ENISA). (2025). ENISA threat landscape 2025. ENISA.

European Commission, Eurostat. (2025). ICT security in enterprises. Statistics Explained. https://ec.europa.eu/eurostat/statistics-explained/index.php?title=ICT_security_in_enterprises

European Commission, Eurostat. (2026). Security incidents and consequences by NACE Rev. 2 activity (isoc_cisce_icn2). Eurostat Database. https://ec.europa.eu/eurostat/databrowser/view/isoc_cisce_icn2/default/table?lang=en

European Commission, Eurostat. (2026). Digital economy and society database. Eurostat. https://ec.europa.eu/eurostat/web/digital-economy-and-society/database

Janikovičová, S., & Barteková, M. (2026). Economic and environmental challenges of the triple transition in small and medium-sized enterprises in the Visegrad Group. Discover Sustainability, 7, 711.

Krstić, J. (2025). Evaluating corporate reputation in the digital environment: Insights from a systematic literature review. In A. Erturk, M. Özşahin, & V. Simovic (Eds.), Human-centric, sustainable, and resilient organizations in the digital age (pp. 235–256). IGI Global Scientific Publishing.

Markovič, P., Pollák, F., Vavrek, R., & Kostiuk, Y. (2022). Impact of coronavirus pandemic on changes in e-consumer behaviour: Empirical analysis of Slovak e-commerce market. Ekonomický časopis, 70(4), 368–389.

Molavi, H., & Zhang, L. (2025). AI-driven corporate reputation measurement in digital ecosystems: A systematic literature review. Acta Psychologica, 254, Article 105846.

Nuortimo, K., Harkkonen, J., & Breznik, K. (2024). Exploring corporate reputation and crisis communication. Journal of Marketing Analytics.

Pollák, F., Dorčák, P., & Markovič, P. (2021). Corporate reputation of family-owned businesses: Parent companies vs. their brands. Information, 12(2), 89.

Rust, R. T., Rand, W., Huang, M.-H., Stephen, A. T., Brooks, G., & Chabuk, T. (2021). Real-time brand reputation tracking using social media. Journal of Marketing, 85(4), 21–43.

Tong, S. C., & Chan, F. F. Y. (2022). Revisiting trust in the digital era: The interplay of online trust and online dialogic communication from the practitioners' perspective. Journal of Communication Management, 26(3), 271–293.

Vassallo, J. P., Seo, Y., & Ansari, S. (2024). Reputation-damaging events over a long time horizon: An event-system model of substantive reputation repair. Journal of Management, 51(3), 1231–1262.

PROJECT MANAGEMENT, STRATEGIC MANAGEMENT AND THEIR INTERCONNECTION IN PRACTICE

Hana Theer Vítková, Hana Mohelská

University of Hradec Kralove
Faculty of Informatics and Management
hana.vitkova@uhk.cz, hana.mohelska@uhk.cz

DOI: 10.35011/IDIMT-2026-425

Keywords

Strategic management; project management; portfolio management; higher education; public administration.

Abstract

The paper addresses the problem of insufficient alignment between strategic management and project management in public administration and higher education institutions. The aim is to identify existing approaches to the use of project management within strategic management and to outline a suitable research design for the effective use of these tools in further empirical investigation. The review shows increasing international emphasis on portfolio management and strategic alignment, while Czech institutions apply these tools inconsistently.

1. Introduction

Project and strategic management both refer to terms such as goals, visions, and ways of achieving them. Therefore, it makes sense to use project management tools for strategic planning, which is becoming increasingly common in the Czech Republic not only in business practice but also in public administration (hereinafter referred to as PA) and, gradually, in setting the strategy of public universities. Strategic management in the context of PA and universities in the Czech Republic is becoming increasingly important. Emphasis is placed on ensuring that the strategy does not remain merely a vision formulated in a document that the institution is required to prepare, but that the set goals are actually achieved within the framework of priorities in all areas that are key to the institution.

2. Methodology

This paper is based on an integrative literature review. Relevant sources were identified in Scopus, Web of Science and Google Scholar using keywords related to strategic management, project management, portfolio management and public administration. Documents were included if they addressed the relationship between strategic and project management or provided methodological frameworks applicable to public institutions and universities. The analysis followed a thematic synthesis approach, comparing international and Czech sources and identifying recurring patterns and gaps. The aim of this literature review was to identify methodologies and verify and summarize

current knowledge in the field of project and strategic management, primarily in the Czech Republic, with the focus on PA and higher education institutions. Important research questions in addressing this topic include:

- RQ1: Are there any established outputs in the literature or practice connecting project and strategic management in business practice?
- RQ2: Are there any known outcomes of connecting project and strategic management in PA, and are there any specific features?
- RQ3: Do universities in the Czech Republic use project management to set their strategy?
- RQ4: What research design can be used in the next phase for a deeper practical integration of project and strategic management, followed by field testing, which will serve as the output of the doctoral thesis?

3. Strategic management process

Strategic management is a continuous, cyclical process involving several interrelated phases. The organization builds on its vision, mission, and purpose, which it specifies in strategic documents. The strategy is implemented through action plans and the setting of key processes, and the implementation is subsequently evaluated. The results of the evaluation may lead to adjustments to the strategic plan, a redefinition of strategic objectives, or a revision of action steps. The entire process is also conditioned by the availability of organizational resources, which determine the realistic possibilities for fulfilling strategic objectives. Based on resource capacities, individual objectives are prioritized, and their implementation is optimized.

The success of strategic management is influenced by several factors. When formulating and implementing a strategy, effective communication at all levels of the organization's management is important, which is closely linked to the selection of competent staff. At the same time, it is necessary to consider the risks that may affect the achievement of the set goals. The risk may be uncertainty from the external environment, but it may also arise within the organization, e.g., if the formulated goals are in conflict.

3.1. Project management

Project management is divided into phases that vary according to individual methodologies and approaches. We typically refer to the pre-project, project, and post-project phases (Pitaš et al., 2012).

3.2. Portfolio management

Portfolio management involves managing selected projects that contribute to fulfilling the vision and strategic development goals of the entity. When managing a project portfolio, it is important to respect the rules of project management, especially limited resources and time constraints. Fotr (2015) states: *"These projects are an important tool for implementing corporate strategy; projects that contribute most to achieving strategic goals and bring the greatest value should be implemented."* Determining the content and scope of the project portfolio and its subsequent management is linked to the strategy of the institution, so the selection of projects for the portfolio should be carried out very carefully. An example is the Bible-Bivins process model, which clearly shows the link between strategic and project management. First, it is necessary to define the vision, strategic goals, and their priorities, and then select the projects from the project plans that best correspond to the fulfilment of the given vision for the portfolio. Evaluation criteria and subsequent criteria for the fulfilment of the strategy according to the project portfolio must be established for all steps.

3.3. Balanced Scorecard

The Balanced Scorecard is a strategic management tool that translates a strategic vision into measurable goals from several perspectives, including financial. It allows you to assess how individual projects contribute to the fulfilment of the strategy, thereby supporting their prioritization. Once strategic goals have been set, key projects are identified and resources are allocated to them according to their strategic relevance.

4. Findings

The literature review focuses on this issue from both an international and a Czech perspective, covering both general aspects and specific applications in public administration and universities.

4.1. Foreign resources

Foreign resources contain articles and studies dealing with the interconnection of project and strategic management in various sectors, including public administration. They unanimously point to how essential it is to connect these two areas in order to fully realize an organization's vision and, in the case of business entities, to increase profits or market position. A number of articles link this call for integration to the COVID-19 pandemic, which has led to changes in the rules set by organizations in many areas.

Bonomi (2024) emphasizes the need to integrate project and strategic management in public administration, where no system or process has been established yet. Therefore, he and his team proposed a process-oriented introduction of project and portfolio management tools and techniques into the strategic planning of public organizations with the aim of improving the quality and efficiency of services.

Gordana Urga (2018), who studied project management in Slovenian public administration, came to similar conclusions. She mentions that public administration faces increasing demands for efficiency, economy, and quality. In order to achieve these changes, it must introduce new operating principles and set a new strategy, which it can anchor through project management. She points out that the strategy is implemented through projects or project portfolios.

Public institutions in South Africa also deal with the integration of these tools in public administration, where Gerrit Van Der Waldt (2016) points out that it is important to link organizations, their management, and strategy with projects so that projects are not implemented randomly and limited resources are not wasted. At the same time, it is important to set up a portfolio of projects based on strategy so that they collectively contribute to fulfilling the vision at all levels of the public sector (Van Der Waldt, 2016).

Hilago (2019) focuses on the principles of strategic management in scientific activities in an academic environment. He uses project management tools to set up strategic management. Using a case study, he examined how the use of project tools, such as mind mapping, brainstorming, stakeholders, decision-making, and the quality of research team collaboration, influences the research process. The use of new methods can improve team cohesion and a shared vision in research (Senabre Hidalgo & Fuster Morell, 2019).

The articles also present opposing theories that incorporate strategic management into project management. One of the authors of such research, Parker et al. (2015), deals with current issues such as resource constraints in his study and uses the so-called triple imperative in his research, i.e., the scope, quality, and time of the required output. His research is based on working with resources and

the impact of their availability and flexibility on achieving the project's goals in the required time and quality. He then illustrates the use of his theory with several practical examples, such as the construction of a tunnel at London Heathrow Airport.

One interesting source that links education, academic activity, and healthcare (an area in which I work) is an article by Leila Maria de Abreu Jaggi from Brazil (2016), which responds to a current topic, namely linking teaching in a professionally oriented study programme and the provision of healthcare services, and it is based on processes defined by project management. It is an educational product that leads to an increase in the professional competence of healthcare personnel, which is one of the strategic goals of the healthcare facility in question, specifically the oncohematology department. It uses the individual stages of project management, namely the planning of the educational product, its integration into the educational process, monitoring the effectiveness of the educational product, modifying the product based on feedback, and finally its definitive integration into the study programme.

Directly from the academic environment comes the output of Yulia Stukalina (2016) presented at a conference in Turkey, which deals with the coordination of international multidisciplinary scientific teams at universities. Internationalization and multidisciplinarity are increasingly important strategic goals for universities because they can enable scientific and research excellence. Generally, a matrix organizational structure is used in such teams. The project outputs are then closely linked to the outputs of the strategy of the partners involved in the team.

Strategic management of universities and cooperation in the field of strategy at the university level are also addressed by project calls within various European programs. Among the most significant is the Horizon Europe call, where current calls focused on strategic management in connection with infrastructure renewal, in the environmental field, and in the field of coordination of the creation and existence of start-up companies can be found. The EIT Higher Education Initiative (2025) call has recently published a pre-call focused on capacity building and strategic cooperation among universities, which is suitable for projects aimed at strengthening strategic skills, entrepreneurship, and university management. From 2022 to the present, an open Erasmus+ call, "Erasmus Charter for Higher Education," has been published with topics supporting cooperation, strengthening university capacities, and management strategies.

4.2. Local resources

The integration of project and strategic management tools in the Czech Republic is primarily an area of focus for the public administration, specifically the Ministry of Regional Development (hereinafter referred to as "MRD"), which deals with the configuration and application of these tools within the public administration. The Ministry has established the "Strategic Work Portal in the Czech Republic" for this purpose (Ministry of Regional Development, n.d.).

The reason for establishing the portal was to fulfil one of the strategic goals focused on client-oriented public administration, for which effectively functioning public administration authorities are a necessary condition.

The MRD presents a diagram on its website (see Figure 1) illustrating strategic working methods using project management. The tools that can also be seen in the setting of university strategic plans include SWOT analysis, risk analysis, project planning and management, and the implementation of public strategies, which mainly builds on the strategy of the Ministry of Education, Youth and Sports, on which university strategic planning is based.

Because of their focus on public administration, these tools can also be used in strategic planning at universities, that is, these documents can be used as a basis for setting university methodology.

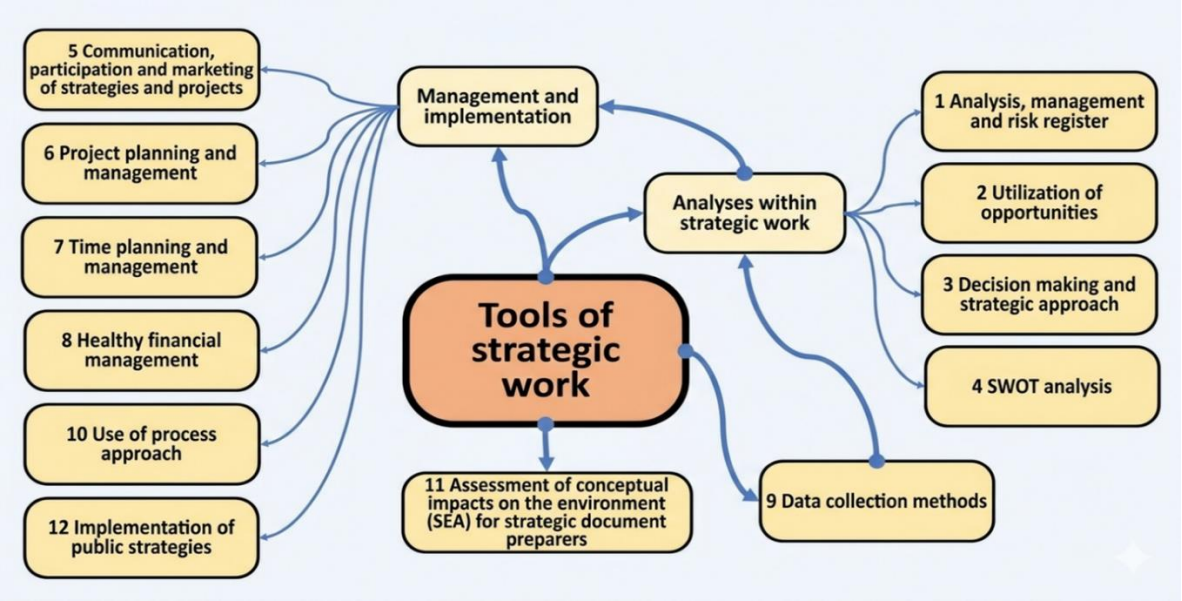


Figure 1. Strategic work tools of the MRD

Source: The output of the project Support for Strategic Management and Planning in Czech Public Administration with an Emphasis on Expanding Methods, Tools, Innovations, Knowledge, and Experience in Practice, registration number CZ.03.4.74/0.0/0.0/15_025/0016924

The Ministry of Education, Youth, and Sports (2021) set out the rules and framework for establishing strategic plans for public universities, which are based on the Strategic Intent for Higher Education and its Implementation Plans. An appendix to this document is the Strategy for the Internationalization of Higher Education for the period starting in 2021, which formulates goals and measures for strengthening the international dimension of university activities. The overarching strategic document that clearly refers to the principles of project management is the Education Policy Strategy of the Czech Republic until 2030+ (Ministry of Education, Youth and Sports, 2020). The document clearly sets out strategic priorities, key activities, monitoring indicators, monitoring and evaluation procedures, a description of management procedures and organizational structure, and a risk management system.

The use of project calls and project tools clearly demonstrates the approach of the Ministry of Education, Youth and Sports, which implemented a project called IPN KREDO (MEYS, 2015) in years 2012 - 2015. The aim of this project was to set out a Strategy for the Development of Higher Education until 2030 in connection with the preparation of strategic plans for the development of higher education institutions until 2030. The project outputs include supporting materials for the preparation of university strategic plans based on background documents, analyses, and studies on key strategic decisions that will influence universities and higher education as a whole in the coming years. Based on these outputs, priorities for higher education are then set, including indicators for monitoring the fulfilment of the visions and objectives of the strategic plans.

All 26 public universities in the Czech Republic prepare strategic documents based on the Strategic Intent for Higher Education and its Implementation Plans (Ministry of Education, Youth and Sports, 2021). At the same time, however, universities participate in calls for development projects primarily from the resources of the Ministry of Education, Youth and Sports' operational programs. The projects focus on the development of strategic management, the implementation of new strategic tools, and, at the same time, internationalization as one of the strategic priorities. The use of project calls to set the strategy shows that universities define the strategy and then set up a specific project as an implementation tool (with its key activities, budget, plan, etc.). There is a clear link between the

university's objectives (strategy) and the project framework (what we will do, when, by whom, with what). Monitoring of the implementation process and the achievement of objectives is then a matter of course.

4.3. Implication for further research

The findings address the research questions by comparing Czech and international approaches (RQ1). While the integration of strategic and project management is considered a standard managerial process in foreign institutions, in Czech institutions, this interconnection is treated primarily on a theoretical level without significant practical application (RQ2). Higher education institutions respond mainly to project calls focused on strategic management, yet the long-term benefit remains limited (RQ3). These findings will be followed by an analysis of university strategic documents, a questionnaire survey, case studies of successful projects, and structured interviews, aiming for a deeper practical integration of project and strategic management with subsequent field testing (RQ4).

5. Discussion

Methodological support from the Ministry of Regional Development and the Ministry of Education, Youth and Sports plays an essential role, providing a framework that must be adapted to the conditions of individual institutions. The success of the strategy therefore depends on the ability to effectively implement these methodologies.

Other barriers include insufficient human resources and poor systematic evaluation of projects. These factors confirm the need to professionalize roles and develop competencies in the area of strategic and project management.

5.1. Recommendation

- Integration of project management into the strategic cycle – organizations should view project management as an integral part of strategic planning, implementation, and evaluation to ensure consistency between strategic goals and their fulfilment.
- High-quality formulation of strategic objectives – a clear and measurable definition of strategic objectives is essential for the effective structuring of project activities. Deficiencies in this area are a frequent problem identified in the document.
- Optimization of organizational structures and communication – clear definition of competencies, responsibilities, and communication processes is key to successful strategy implementation, as it eliminates conflicts of priorities and improves coordination.
- Development of portfolio management – projects should be selected based on their strategic relevance, not just the availability of funding. Portfolio management promotes the optimal use of limited resources while ensuring synergy across individual projects.
- Support for the professional development of employees – a proposal for systematic training in project management and strategic planning. Human resources are a significant determinant of the successful implementation of the strategy.
- Use of new modern software tools with the potential for AI integration – the implementation of specialized tools for project and portfolio management promotes transparency and accuracy in planning processes, which can significantly improve the management of strategic activities.
- Involvement in international projects – participation in international programs such as Horizon Europe or Erasmus+ strengthens the strategic competencies of institutions, their scientific profile, and their capacity for international cooperation.

5.2. Research limitations

The research reveals several limitations that may affect the interpretation of the conclusions:

- Limited availability of documents: some institutions do not publish detailed internal procedures, which may reduce the completeness of the analysis.
- Variability of terminology: differences in the definitions of strategic and project management in professional literature can complicate direct comparisons of sources.
- Focus on public universities: the transferability of conclusions to other sectors may be limited due to differences in management and organizational culture.
- Absence of empirical research: this phase of the research is based on literary analysis; empirical data will be obtained in the subsequent part of the research through interviews and benchmarking to be carried out via an examination of strategic materials and annual reports (RQ4).
- Dynamics of strategic documents: strategic materials of institutions are regularly updated, and therefore some conclusions may reflect the situation valid only at the time of the research.

6. Conclusion

This literature review confirms that strategic management and project management form an interconnected system, and that the integration of these two systems is essential for the effective functioning of organizations. Project management makes it possible to translate strategic objectives into specific, measurable, and controllable activities, thereby supporting their implementation in an environment with limited resources and growing demands for transparency and efficiency. This interconnection is evident in professional literature and in the methodological frameworks of public administration authorities that support systematic project planning and project management.

Problems in linking strategic and project management can include poorly defined and formulated strategic goals and visions and ignoring limited financial resources. Poorly designed organizational structures can also prevent the implementation of strategies, leading to poor communication and coordination. Conflicts of priorities can arise at various levels of the hierarchical structure. The established visions then fail to consider the need for human resources with the necessary knowledge and skills.

6.1. Software tools for strategic planning

There are a number of software tools available for the proper configuration of projects, enabling the management of project portfolios and strategic objectives. Abroad, large universities use Planview software, which focuses on strategic planning and performance measurement. The public administration sector tends to use SAP Project System software, which allows strategies to be linked to financial resources. In the Czech Republic, the public administration sector and universities still most commonly use Microsoft Project with Gantt charts, resource management, and project portfolios.

6.2. Internationalization in project and strategic management

The above-mentioned background information and conclusions clearly demonstrate the necessity and importance of internationalization in the field of strategic management. Outside the Czech Republic, the interconnection of strategic and project management is becoming commonplace in the private sector, and it is also being used in academic environment. The internationalization is one of the key parameters of strategic planning for Czech public universities. It is beginning to have a significant

impact in the area of scientific and research cooperation. International project calls, particularly from the Horizon Europe program, also focus on university strategy and cooperation in this area. Another important project and financial tool that is relatively widespread is the involvement of domestic universities in the Erasmus+ program, in particular by establishing cooperation with foreign universities inside and outside the EU in the field of education, sharing teaching experiences, setting up new teaching methods, and standardizing professional education in certain areas.

Acknowledgements

This work was supported by the internal project “SPEV 2026, University of Hradec Králové, Faculty of Informatics and Management, Czech Republic. "Strategic Approaches to Knowledge Management in Large Multinational Corporations", principal investigator: Chinyere Chidera Okechukwu, MBA.

Resources

Bonomi Savignon, A., & Costumato, L. (2024). Project management logics for agile public strategic management: Propositions from the literature and a research agenda. Information Polity, 29(2), 153–178. https://doi.org/10.3233/IP-230061

EIT Higher Education Initiative. (2025). EIT Higher Education Initiative launches €70 million call to boost STEM innovation and strengthen university cooperation. https://eit-hei.eu/eit-higher-education-initiative-launches-70-million-call-to-boost-stem-innovation-and-strengthen-university-cooperation/

Fotr, J., & Souček, I. (2015). Tvorba a řízení portfolia projektů: Jak optimalizovat, řídit a implementovat investiční a výzkumný program. Grada. https://www.bookport.cz/kniha/tvorba-a-rizeni-portfolia-projektu-2420/

Ministerstvo pro místní rozvoj. Portál strategické práce v České republice. https://mmr.gov.cz/cs/microsites/portal-strategicke-prace-v-ceske-republice/strategicke-projekty/podpora-strategickeho-rizeni-ve-verejne-sprave-cr

Ministerstvo pro místní rozvoj. Nástroje strategické práce. Portál strategické práce v České republice. Ministerstvo pro místní rozvoj ČR - Nástroje pro strategickou práci

Ministerstvo školství, mládeže a tělovýchovy. (2015). IPN KREDO: Kvalita, relevance, efektivita, diverzifikace a otevřenost vysokého školství v ČR. Strategie vysokého školství do roku 2020 v návaznosti na Evropu 2020 a nový zákon o vysokých školách. https://msmt.gov.cz/vzdelavani/vysoke-skolstvi/ipn-kredo-kvalita-relevance-efektivita-diverzifikace-a

Ministerstvo školství, mládeže a tělovýchovy. (2020). Strategie vzdělávací politiky ČR do roku 2030+. https://msmt.gov.cz/vzdelavani/skolstvi-v-cr/strategie-2030

Ministerstvo školství, mládeže a tělovýchovy. (2021). Strategický záměr pro oblast vysokých škol a plány jeho realizace. https://msmt.gov.cz/vzdelavani/vysoke-skolstvi/strategicky-zamer

Parker, D. W., Parsons, N., & Isharyanto, F. (2015). Inclusion of strategic management theories to project management. International Journal of Managing Projects in Business, 8(3), 552–573. https://doi.org/10.1108/IJMPB-11-2014-0079

Pitaš, J., Hajkr, J., Havlík, J., Máchal, P., Motal, M., et al. (2012). Národní standard kompetencí projektového řízení verze 3.2. Společnost pro projektové řízení, o. s.

Senabre Hidalgo, E., & Fuster Morell, M. (2019). Co-designed strategic planning and agile project management in academia: case study of an action research group. Palgrave Communications, 5(1), Article 151. https://doi.org/10.1057/s41599-019-0364-0

Vacík, E., Špaček, M., Fotr, J., & Kracík, L. (2018). Project portfolio optimization as a part of strategy implementation process in small and medium-sized enterprises: a methodology of the selection of projects with the aim to balance strategy, risk and performance. E+M Ekonomie a Management, 21(3), 107–123. https://doi.org/10.15240/tul/001/2018-3-007

Van Der Waldt, G. (2016). Probing strategy-project alignment: The case of the South African Social Security Agency. The Journal for Transdisciplinary Research in Southern Africa, 12(1), a358. https://doi.org/10.4102/td.v12i1.358

ANNEX

STATEMENT OF THE PUBLICATION ETHICS AND PUBLICATION MALPRACTICE

IDIMT's Publication Ethics and Publication Malpractice Statement is based, in large part, on the guidelines and standards developed by the Committee on Publication Ethics (COPE).

We expect all parties commit to these publication ethics. We do not tolerate plagiarism or other unethical behaviour and will remove any manuscript that does not meet these standards.

The relevant duties and expectations of authors, reviewers, and editors are set out below:

1. Author Responsibilities

Authors must certify that their manuscripts are their original work, and they take full responsibility for it, including parts created/based on AI.

Authors must certify that the manuscript has not previously been published elsewhere.

Authors must certify that the manuscript is not currently being considered for publication elsewhere.

Authors must notify us of any conflicts of interest.

Authors must identify all sources used in the creation of their manuscript.

Authors must report any errors they discover in their manuscript.

Authors must list all persons who contributed to the manuscript (co-authors, collaborations, etc). AI systems cannot be named as (co-)authors

Authors must identify all parts created by or substantially based on AI tools (e.g., created/rewritten by LLMs, translated by software). The model or tool used and its manufacturer must be cited.

Authors remain fully responsible for the content of their manuscripts, including ensuring the accuracy of the information and maintaining the originality and integrity of their work.

2. Reviewer Responsibilities

Reviewers must notify us of any conflicts of interest.

Reviewers must keep information pertaining to the manuscript confidential.

Reviewers must bring to the attention of the Editor-in-Chief any information that may be reason to reject publication of a manuscript.

Reviewers must at any time evaluate manuscripts only for their intellectual content without regard to race, gender, sexual orientation, religious belief, ethnic origin, citizenship, or political philosophy of the authors.

Reviewer who feels unqualified to review the research reported in a manuscript or knows that its prompt review will be impossible should notify us and excuse himself from the review process.

3. Editorial Board Responsibilities

The Editorial Board must keep information pertaining to submitted manuscripts confidential.

The Editorial Board must disclose any conflicts of interest.

The Editorial Board must evaluate manuscripts only for their intellectual content.

The Editorial Board is responsible for making publication decisions for submitted manuscripts.

LIST OF AUTHORS

A

Aggarwal Shalini 167, 259
Ahmed Sharfraj..... 259
Al Sardy Loui 135
Anderle Michael 197
Anwar Talha 33
Ardielli Eva..... 99

B

Bahardoust Selina 15
Bachura Martin 267
Baklizi Mahmoud 135
Barkóciová Miroslava..... 275
Bartekova Maria 107
Baše Radan 341
Bocánová Slávka 267
Boltun Erik 167, 323
Borstnar Mirjana Kljajic 91
Brontvaj Marián..... 283

Č

Čuka Ondřej..... 341

D

Delina Radoslav..... 247
Demeter Gabriel 299
Dhawan Prerna 167, 259
Dolná Mária..... 291
Doucek Petr 155
Džujková Alžbeta 331

E

Espinoza Felix 155

F

Feranecová Adela 359
Freňáková Magdaléna.... 267, 291

G

Gocze Abraham 299
Grima Simon..... 167, 259

H

Hagendorn Melissa 15
Herout Lukas 205
Herzigová Kristína..... 383
Hvastová Jozefína..... 291

I

Ignjatovic Drazen..... 15
Ivaničková Mariana 275

J

Janošková Mária.....351
Jaroš Miroslav 341
Jiříčková Ludmila.....401
Jung Martin 33
Jurovčáková Tatiana.....267

K

Kašpar Vladislav 307
Katrenčik Ivan.....81
Klarner Lukáš.....25, 409
Koba Patrik 247, 375
Králová Petra.....367
Kratky Daniel 315
Krusiec Krzysztof.....63
Kumar A Sujith 167

L

Leidwein Alois.....15
Lisnik Anton 55
Lukáš Martin 197

M

Maryška Miloš 155
Maříková Monika.....25
Matoušková Petra.....25
Mohelská Hana.....425
Muhammad Mamdouh 135

N

Němečková Kamila.....25, 409
Neubauer Georg 15, 33
Novak Richard A.....185
Nováková Alena.....359

O

Olexa Natália.....383
Olexová Cecília.....375, 383

P

Palej Andrzej.....63
Parzer Verena.....33
Pavlíček Antonín....185, 221, 229
Pitner Tomáš 341
Pollák František.....393, 417
Pucihar Andreja.....91

R

Rainer Karin.....15
Rao Jennifer 15
Reichl Bernd..... 33
Rolínek Ladislav..... 25
Rožmberská Lucie 213

Ř

Řehoř Petr..... 25, 409

S

Shaaban Abdelkader Magdy. 135, 143
Schmittner Christoph 143
Schoitsch Erwin..... 127
Sieber Jakub..... 323
Sigmund Tomáš..... 73
Skřehota Ondřej..... 341
Sobotová Romana..... 417
Sonntag Michael 43
Sood Kiran..... 167, 259
Soukupová Nikola 409
Sudzina František..... 185, 221
Sulová Michaela 247
Sumereder Anna 15
Syrovátková Jana 185, 229

Š

Štěpánek Lubomír..... 237

T

Tkáč Michal..... 331
Tkáč Michal..... 331
Trenz Oldřich..... 341

U

Ulrichova Anna..... 205

V

Veber Jaromír 367
Verner Robert 283, 315
Vítková Theer Hana..... 425
Vrchota Jaroslav 25

W

Weiss Erik..... 259, 351
Weiss Roland..... 167

Z

Zatrochová Monika..... 81
Zbořil Martin 115
Zelenka Martin 175
Zuzik Jozef 351

TRAUNER VERLAG
UNIVERSITÄT

IDIMT-2026
Resilience: ICT, Business and Society

34th Interdisciplinary Information Management Talks

This book presents the proceedings of the 34th annual IDIMT conference.

Past experience has shown that resilience is essential, yet often overlooked. Even a minor issue can cause major disruption when only the "normal" or expected case is considered. This applies to software development (security), AI (reliability of outputs), and the design and use of systems in business and society, including marketing, social media, supply chains and autonomous systems. We examine resilience from technical, ethical and management perspectives, discussing its role in the design, implementation, transformation and use of ICT systems across various contexts, particularly crisis scenarios where resilience is indispensable.

We have chosen the following 11 topics for 2026:

- Early Career & Student Showcase (Sonntag)
- New Paths to tackle Complex Crisis: Decision support by AI and decentralized data management (Neubauer, Rainer)
- Cyberresilience for Security (Sonntag)
- Multiple ethical aspects and limits of ICT resilience (Lisnik)
- Data, AI and Digital-Driven Transformation: Shaping Sustainable Digital Futures (Pucihar)
- Resilience for Autonomous Systems (Schoitsch)
- Data Governance, Data Analytics & AI (Doucek)
- Social Media, Creativity and Innovation: Synergies and Tensions (Pavlíček)
- AI & Data-Driven Supply Chain Management (Delina, Tkáč)
- From Campus to Company: Collaborative Innovation in the Age of AI (Pitner)
- Business, Trade and Marketing in the Age of AI (Janošková)

Based on a double-blind, two-step review process, we selected 47 of the submitted papers, including more than 103 co-authors. The program was organized into two parallel streams.

The authors come from 8 countries: Austria, Czech Republic, Germany, India, Jordan, Malta, Slovakia and Slovenia.

DOI:10.35011/IDIMT-2026

9 783991 952183

ISBN 978-3-99195-218-3

www.trauner.at