Balkans Journal of Emerging Trends in Social Sciences

– Balkans JETSS –

Vol. 8 – No. 2 – 2025

Contents

Balkans Journal of Emerging Trends in Social Sciences

ISSN: 2620-164X

Balkans Journal of Emerging Trends in Social Sciences – Balkans JETSS

Aims and Scope

The mission of Balkans JETSS is to publish peer-review empirical research papers that test, extend or build theory and contribute to practice. All empirical methods – including, but not limited to, qualitative, quantitative, field, laboratory, and combination methods are welcome. Empirical, theoretical and methodological articles from all major fields of economics, management, tourism, law and the like are featured in the journal. Theoretical and/or review articles that integrate existing bodies of research and that provide new insights into the field are also encouraged.

To be published in the Balkans JETSS, a manuscript must take strong empirical and/or theoretical contributions to the subject field. Consequently, preference is given to submissions that test, extend or build strong theoretical frameworks while empirically examining issues with high importance for theory and practice.

The journal is not tied to any particular discipline, level of analysis, or national context. Although, it focuses on Balkans region, all papers from related fields on any region or country are highly encouraged. Single country studies, multi-country or regional studies can be submitted.

Manuscripts should not exceed 16 pages (450 word per page). This page limit includes all figures, tables, appendices and references.

Copyright information

Instructions for manuscripts

Manuscripts should be in English. Under the paper title, the name(s) of the author(s) should be given. Affiliation should be placed in the footnote together with the exact mail and e-mail address.

Manuscript format. A brief abstract of approximately 100 to 150 words and a list of up to six key words should precede the text body of the manuscript. Also, an appropriate number of JEL codes should be provided. This classification system is prepared and published by the *Journal of Economic Literature*, see www.aeaweb.org/journal/jel_class_system.html.

Manuscripts should be prepared as doc file, Word version 6.0 or higher.

Manuscript length. Brief articles and discussions (10 pages or less - 450 words per page) are encouraged, otherwise, papers should present well-focused arguments of approximately 16 pages.

Style requirements. Letters, figures and symbols should be clearly denoted.

Equations should be typewritten and with the number places in parenthesis at the right margin. References to equations should be in the form “Eq, (2)” or simply (2). For equations that cannot be entered in a single line, use the Equation Editor in MS Word. In equations and in the text italicize symbols that are used to represent variables or parameters, including subscripts and superscripts. Only use characters and symbols that are available in the *Equation Editor*, in the *Symbol font* or in the *Times New Roman*.

All illustrations (figures, photographs, line drawings, graphs) should be numbered in series and all legends should be included at the bottom of each illustration. All figures, photographs, line drawings and graphs, should be prepared in electronic form and converted in TIFF of JPG (max quality) file types, in 300 dpi resolution, for superior reproduction. Figures, line drawings and graphs prepared using elements of MS Drawing of MS Graph must be converted in form of pictures and unchangeable. All illustrations should be planned in advance so as to allow reduction to 12.75 cm in column width. Please review all illustrations to ensure that they are readable.

All *tables* should be numbered with consecutive Arabic numbers. They should have descriptive captions at the top of each table and should be mentioned in the text.

References should follow the APA Style convention, in alphabetical order at the end of the manuscript. The list of references should be formatted so that the second row in each entry in indented (paragraph indentation, special – choose hanging, by 0.5 cm). Wherever possible, the DOI number should be provided, too, in addition to other reference data.

The APA style citation is applied in the text (according to the instructions that can be downloaded from the link http://www.apastyle.org). Citations in the text should be given in brackets stating the author's surname, year of publication and, possible, pages, if it is a direct quote).

The authors themselves are responsible for the correctness of the English language in papers.

Electronic submission

Papers for consideration should be submitted to the Balkans JETSS editor in electronic form via journal management and publishing software at the https://balkans-jetss.org

Managing Editor
Nikolina Vrcelj, PhD
Association of Economists
and Managers of the Balkans,
13/2 Džona Kenedija St, 11070 Belgrade, Serbia
Phone +381 62 812 5779
E-mail: nvrcelj@balkans-jetss.org
https://www.balkans-jetss.org

Technical Editor
Branimir Trošić

Publication frequency – one volume, two issues per year.

Publisher – Association of Economists and Managers of the Balkans, Belgrade, Serbia.

Co-publisher – Faculty of Tourism and Hospitality – Ohrid, University "St. Kliment Ohridski" – Bitola, Ohrid, North Macedonia.

Printed by Skripta Internacional, Belgrade, Serbia

Balkans Journal of Emerging Trends in Social Sciences uses the **Crossref Similarity Check powered by iThenticate** software to detect instances of overlapping and similar text in submitted manuscripts. All can be reassured that the publisher's content is committed to actively combating plagiarism and publishing original scientific research.

Disclaimer: The author(s) of each article appearing in this Journal is/are solely responsible for the content thereof; the findings, interpretations and conclusions expressed in the articles are those of the authors and do not reflect the view of the Editors, Reviewers, the Publisher or anyone else involved in creating, producing or delivering the Balkans Journal of Emerging Trends in Social Sciences.

Issue DOI:
https://doi.org/10.31410/Balkans.JETSS.2025.8.2

CIP – Katalogizacija u publikaciji
Narodna biblioteka Srbije, Beograd
316.42
BALKANS Journal of Emerging Trends in Social Sciences : Balkans JETSS / editor-in-chief Aleksandar Trajkov. - Vol. 1, no. 1 (2018)- . - Belgrade : Association of Economists and Managers of the Balkans, 2018- (Belgrade : Skripta Internacional). - 25 cm
Polugodišnje.
Dostupno i na: http://www.udekom.org.rs/balkans-jetss.html .
ISSN 2620-164X = Balkans Journal of Emerging Trends in Social Sciences
COBISS.SR-ID 270984460

List of the Reviewers for 2025

Balkans Journal of Emerging Trends in Social Sciences wishes to thank the reviewers over the past year who have performed an essential role in maintaining the quality of this publication. Their comments and suggestions were of great help to the authors in improving the quality of their papers. Each of the reviewers listed below returned at least one review in 2025.

Prof. **Ali Maksuti**, PhD
University of Tetovo, Str. Ilinden No. Tetovo, North Macedonia
Prof. **Anton Vorina**, PhD
Celje School of Economics: Celje, Slovenia
Prof. **Beatrix Séllei**, PhD
Budapest University of Technology and Economics, 1117 Budapest, Magyar tudósok körútja 2., Hungary
Prof. **Beatriz Corchuelo Martínez-Azúa**, PhD
University of Extremadura, Avda. de Elvas, s/n. 06006 Badajoz, Spain
Prof. **Brikene Dionizi**, PhD
University of Shkoder "Luigj Gurakuqi", Sheshi 2 Prill, Shkoder, 4001, Albania
Prof. **Carmen Nastase**, PhD
Faculty of Economics, Administration and Business, Stefan cel Mare University, University Street, 13, 720229, Suceava, Romania
Prof. **Christian Tanushev**, PhD
University of National and World Economy, Student Town, Sofia 1170, Bulgaria
Prof. **Daniela Mechkaroska**, PhD
University of Information Science and Technology "St. Paul the Apostle" - Ohrid, Macedonia, Partizanska bb, 6000 Ohrid
Dijana Štrbac, PhD
Institute of Economic Sciences, Belgrade, Zmaj Jovina 12, Serbia
Prof. **Dunja Mestrovic**, PhD,
Centre for Quality Assurance and Institutional Research at the University of Rijeka, Croatia
Prof. **Dženis Bajramović**, PhD
University of Novi Pazar, Dimitrija Tucovića bb, Novi Pazar, Serbia
Prof. **Eka Lekashvili**, PhD
Ivane Javakhishvili Tbilisi State University, Ave. Tchavtchavadze, 1. Tbilisi, Georgia
Prof. **Emilija Gjorgjioska**, PhD
Faculty of Economics Prilep, University St Kliment Ohridski Bitola, Prilepski Braniteli br 143, North Macedonia
Prof. **Entela Nikaj**, PhD
Law, University of Tirana, Rruga Milto Tutulani, Tirana, Albania
Prof. **Irina Teodora Manolescu**, PhD
Alexandru Ioan Cuza University of Iasi, 11 Bd. Carol I, Iasi, Romania
Prof. **Ivana Ercegovac**, PhD
Higher Colleges of Technology, Kuwait Rd., Al Faseel, Fujairah, UAE
Prof. **Jakub Sopko**, PhD
Faculty of Economics, Technical University of Košice, Němcovej 32, 040 01 Košice, Slovakia
Prof. **Jelena Blaži**, PhD
University North, Jurja Križanića 33, Varaždin, Croatia
Prof. **Jose Manuel Argudin Bringas**, PhD
Universidad Autonoma De Madrid, Iudad Universitaria Cantoblanco, Spain
Prof. **Julianna Csugány**, PhD
Eszterházy Károly Catholic University, 3300, Eger, Egészségház utca 4., Hungary
Prof. **Koviljka Banjević**, PhD
Academy of Applied Technical Studies Belgrade, Department of Belgrade Polytechnic, Katarine Ambrozic 3, Belgrade, Serbia
Prof. **Leoš Šafár**, PhD
Faculty of Economics, Technical University of Košice, Němcovej 32, 040 01 Košice, Slovakia
Prof. **Linda Martic Kuran**, PhD
Polytechnic Marko Marulic Knin, Petra Kresimira 30, Knin, Croatia
Prof. **Marija Ivaniš**, PhD
Faculty of Tourism and Hospitality Management, University of Rijeka, Croatia
Prof. **Matea Matić Šošić**, PhD University of Dubrovnik, Faculty of Economics and Business, Lapadska obala 7, 20000 Dubrovnik, Croatia
Prof. **Michal Mešťan**, PhD
Faculty of Economics, Matej Bel University in Banská Bystrica, Tajovského 10, 975 90 Banská Bystrica, Slovakia

Balkans JETSS (2025) 2: 69-79
DOI: https://doi.org/10.31410/Balkans.JETSS.2025.8.2.69-79

Original Scientific Article

Artificial Intelligence in Serbian Enterprises: Adoption Levels, Obstacles and Sectoral Applications

Lazar Živković[1]
Dijana Štrbac[2]
Nikola Vasilić[3]

Received: October 10, 2025 / Accepted: December 30, 2025 / Published: December 30, 2025

Abstract: *This paper examines the adoption of artificial intelligence (AI) among Serbian enterprises, with a focus on adoption levels, obstacles, and sectoral applications. Based on a survey of 499 companies conducted in 2025, the findings show that only 29.1% of firms use AI technologies, with adoption marginally higher in manufacturing (31.1%) than in services (29.2%), although the difference is not statistically significant. The main barrier to adoption is the perception that AI is not useful or relevant for business operations, reported by 66.9% of non-adopters, followed by concerns about legal uncertainty, lack of knowledge, and high costs. Among adopters, the most frequent applications are data analysis and interpretation, customer support, and personalization in marketing. Sectoral differences are evident: services focus on customer support, while manufacturing emphasizes personalization and product development. The results suggest that informational barriers and limited internal capacities outweigh purely technological constraints, highlighting the need for targeted policy measures to foster AI adoption.*

Keywords: *Artificial intelligence, Adoption barriers, Manufacturing, Services, Serbia*

JEL Classification: O33 · L25

✉ lazar.zivkovic@ien.bg.ac.rs
[1] Institute of Economic Sciences, Belgrade, Zmaj Jovina 12, Belgrade, Serbia
[2] Institute of Economic Sciences, Belgrade, Zmaj Jovina 12, Belgrade, Serbia
[3] Institute of Economic Sciences, Belgrade, Zmaj Jovina 12, Belgrade, Serbia

1. INTRODUCTION

Artificial intelligence (AI) has emerged as one of the most transformative technologies of the 21st century, reshaping business models, organisational processes, and competitive dynamics across industries. As AI-driven solutions are increasingly integrated into data analytics, production optimisation, marketing, and customer interaction, enterprises worldwide recognise AI as a strategic asset that enhances innovation, efficiency, and long-term competitiveness (Czarnitzki et al., 2023; Kassa & Worku, 2025; Li et al., 2023; Wang & Liu, 2025). In parallel, governments and international organisations emphasise AI as a key pillar of digital transformation and industrial modernisation, reflecting its growing economic and societal relevance (European Commission, 2020; OECD, 2025; World Economic Forum, 2025).

In line with these global developments, Serbia has positioned artificial intelligence as a national strategic priority. In January 2025, the Government of Serbia adopted a new Strategy for the Development of Artificial Intelligence for the period 2025-2030, defining objectives related to legal and ethical frameworks, infrastructure development, education, and the deployment of AI systems in the public sector. This strategy builds on the earlier 2020-2025 AI framework, under which Serbia became the first country in Southeast Europe to adopt a national AI strategy, and on the establishment of a National AI Platform within the State Data Center in Kragujevac. Together, these initiatives aim to create favourable conditions for AI diffusion across the economy.

These policy developments are particularly relevant for small and medium-sized enterprises (SMEs), which constitute the backbone of the Serbian economy. For SMEs, AI has the potential to support digital transformation, improve operational efficiency, and enhance competitiveness. However, the pace and depth of AI adoption vary considerably across countries, sectors, and firm sizes. While large firms in advanced economies have increasingly embedded AI into core business functions, SMEs in emerging and transition economies often lag due to limited resources, skills, and awareness of AI's potential business value (Cannavale et al., 2025; OECD/BCG/INSEAD, 2025). In such contexts, barriers to adoption are frequently organisational or perceptual rather than purely technological

Despite growing policy attention, systematic firm-level evidence on AI adoption in the Western Balkans, and particularly in Serbia, remains limited. Existing empirical studies primarily address broader digitalisation or innovation activities, without focusing explicitly on AI adoption, perceived barriers, or sectoral patterns of use. This gap is particularly important given the structural significance of manufacturing and service sectors in Serbia and their differing innovation and digitalisation profiles.

This study addresses this gap by providing empirical evidence on AI adoption among Serbian enterprises. The analysis focuses on three interrelated dimensions: (1) the level of AI adoption, (2) perceived barriers to adoption, and (3) areas of AI application. Particular attention is paid to differences between manufacturing and service-sector enterprises, as these sectors represent distinct but comparable contexts for AI adoption within the same institutional environment. By doing so, the paper contributes to the understanding of how firms in a transition economy perceive, adopt, and apply AI technologies, offering insights relevant for both researchers and policymakers.

The remainder of the paper is structured as follows. Section 2 reviews the relevant literature on artificial intelligence adoption and firm-level barriers. Section 3 outlines the methodology, including the survey design, sample characteristics, and analytical approach. Section 4 presents and discusses the empirical results. Section 5 outlines directions for future research, and the final section concludes the paper.

2. LITERATURE REVIEW

This study builds on the Technology-Organization-Environment (TOE) framework, which is widely used to explain firm-level adoption of new technologies. According to the TOE framework, adoption decisions are shaped by technological characteristics, organisational capabilities, and environmental conditions (Tornatzky & Fleischer, 1990). Recent research applies this framework to artificial intelligence adoption, particularly in the context of SMEs, highlighting the importance of perceived usefulness, internal knowledge and skills, and regulatory uncertainty (Ghani et al., 2022; Sanchez et al., 2025).

At the technological level, AI is widely recognised for its potential to automate processes, enhance data analysis, and improve decision-making. Empirical studies show that AI adoption can reduce errors, optimise workflows, and improve service quality, particularly in data-intensive and digitally mediated business environments (Kuzminov, 2024; Nesterova, 2024). However, perceived relevance and applicability to specific business models remain critical determinants of whether firms adopt AI in practice.

From an organisational perspective, AI adoption is strongly influenced by firm size, internal capabilities, and managerial orientation. Evidence suggests that large enterprises and knowledge-intensive firms lead AI adoption, while SMEs often lag due to limited financial and human resources and the absence of clear strategic frameworks (Enshassi et al., 2024; Gładysz et al., 2023; OECD/BCG/INSEAD, 2025). Organisational culture also plays a central role: leadership awareness and a data-driven mindset are consistently identified as key enablers of successful AI integration (Oyekunle & Boohene, 2024).

Environmental factors further shape AI adoption decisions. Regulatory uncertainty, ethical concerns, data protection requirements, and the availability and quality of data can significantly affect firms' willingness to adopt AI technologies. These constraints are particularly relevant for SMEs operating in less developed or transition economies, where institutional frameworks and support mechanisms are still evolving.

The literature also points to systematic differences between manufacturing and service sectors in AI adoption. Manufacturing firms tend to adopt AI primarily for process optimisation, quality control, predictive maintenance, and product development, reflecting their capital-intensive and process-oriented nature (Chen & Jin, 2023; Pariso et al., 2025). In contrast, service-sector firms more frequently apply AI in customer-facing activities such as customer support, marketing personalisation, and data analytics, where interaction with customers and digital data is central (Wıjayatı et al., 2022; Xue et al., 2022). These differences suggest that sectoral characteristics shape not only adoption rates but also the types of AI applications pursued by firms.

Despite the growing international literature on AI adoption, empirical evidence from smaller and transition economies remains limited. In the case of Serbia, existing studies largely address digitalisation or innovation more broadly, without systematically examining AI-specific adoption, perceived barriers, or sectoral patterns at the firm level. This gap limits understanding of how enterprises in such contexts perceive and integrate AI technologies.

To address this gap, the present study examines AI adoption among Serbian enterprises, with a particular focus on sectoral differences between manufacturing and services. The analysis is guided by the following research questions:

- RQ1: What is the level of artificial intelligence adoption among Serbian enterprises?
- RQ2: Do AI adoption rates differ between manufacturing and service-sector enterprises?
- RQ3: What are the main barriers preventing Serbian enterprises from adopting AI technologies?
- RQ4: In which business functions are AI technologies most commonly applied, and do these patterns differ across sectors?

3. METHODOLOGY

The study relies on primary firm-level survey data collected by the Institute of Economic Sciences in Serbia in May 2025 as part of a broader research initiative examining innovation and innovation-related activities among SMEs. The survey was designed to capture firms' engagement in innovation processes, digital transformation, and the adoption of advanced technologies, including artificial intelligence. The target population consisted exclusively of enterprises with more than nine employees; micro-enterprises were therefore intentionally excluded from the sample. This design choice reflects the analytical focus on firms possessing a minimum level of organisational structure and internal capacity relevant for innovation activities and technology adoption.

The sampling framework was constructed to ensure representativeness at both the sectoral and regional levels. Specifically, the sample reflects the distribution of economic activity across NACE Rev. 2 sectors, as well as the territorial structure of the Serbian economy. Data were collected using a structured questionnaire administered in an online format. Participation in the survey was voluntary, and respondents were assured anonymity in order to promote accurate and unbiased reporting. Within each enterprise, the questionnaire was addressed to respondents occupying managerial or decision-making positions relevant to business strategy, innovation, digitalisation, or technology adoption (such as owners, directors, senior managers, or heads of relevant innovation-related departments).

The total survey sample comprises 499 valid responses from enterprises operating in Serbia. For the purposes of this paper, the analysis focuses exclusively on firms in the manufacturing and services sectors, as these sectors are directly comparable in terms of business processes and digitalisation and are particularly relevant for analysing enterprise-level AI adoption. Manufacturing and service enterprises differ in their core processes and data practices, which are central to the feasibility and use of AI technologies. Excluding firms from other sectors enhances comparability and analytical clarity. As a result, the final analytical sample consists of 440 enterprises, including manufacturing firms (N = 180) and service firms (N = 260).

The questionnaire consisted of several thematic sections addressing firm characteristics, digitalisation practices, and the use of advanced technologies. In this paper, the analysis focuses on three groups of variables related to artificial intelligence. First, AI adoption was measured using a binary indicator capturing whether the enterprise uses artificial intelligence technologies in its business operations (yes/no). Second, firms that reported not using AI were asked to indicate the main reasons for non-adoption, selecting from a predefined list of potential barriers, including perceived irrelevance for business operations, legal uncertainty, lack of knowledge, high costs, data-related constraints, ethical concerns, and organisational limitations. Multiple responses were allowed. Third, firms that reported using AI were asked to identify specific areas of AI application, such as data analysis and interpretation, customer support, marketing personalisation, automation of internal processes, forecasting, product or service development, and risk or problem detection. The questionnaire items used to measure AI adoption, perceived barriers to adoption, and areas of AI application are reported in Appendix A.

Sectoral differences in AI adoption were assessed using a chi-square test of independence with Yates' continuity correction. The analysis was conducted on the analytical sample of manufacturing and service enterprises (N = 440).

4. RESULTS AND DISCUSSION

The results show that the overall level of AI adoption remains limited: only 29.1% of firms report using AI technologies, while the majority (70.9%) have not integrated AI into their business processes. Adoption is slightly higher in manufacturing (31.1%) than in services (29.2%), indicating a modest difference in technology uptake (Table 1). However, this difference is not statistically significant. A chi-square test of independence with Yates' continuity correction indicates no association between sector and AI adoption ($\chi^2(1, N = 440) = 0.16$, $p = 0.68$), suggesting that AI uptake is broadly similar across manufacturing and service firms (Table 2).

Table 1. Adoption of AI among Serbian companies by sector

AI Adoption	Total (N=499)	Manufacturing (N=180)	Services (N=260)
Yes	145 (29.1%)	56 (31.1%)	76 (29.2%)
No	354 (70.9%)	124 (68.9%)	184 (70.8%)

Source: Own research

Table 2. Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	.261[a]	1	.609		
Continuity Correction[b]	.164	1	.686		
Likelihood Ratio	.260	1	.610		
Fisher's Exact Test				.672	.342
Linear-by-Linear Association	.260	1	.610		
N of Valid Cases	440				
a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 53.59.					
b. Computed only for a 2x2 table					

Source: Own research

The main reasons for not using AI are the perception that these technologies are not useful or relevant to business models (66.9% overall), followed by concerns related to legal consequences (10.7%), lack of knowledge (9.6%), and high costs (8.8%). These findings suggest that barriers are not primarily technical but rather linked to awareness of potential benefits and the limited availability of internal expertise. Both manufacturing (61.3%) and services (71.2%) predominantly cite the irrelevance of AI as the key reason for non-adoption, pointing to a widespread perception of limited applicability across sectors (Table 3).

This dominance of perceptual barriers in the Serbian context is consistent with international evidence, which highlights that organizational readiness and managerial perceptions of AI's business value often outweigh purely technical constraints. Sanchez et al. (2025) demonstrate that in SMEs, adoption decisions are frequently shaped by a lack of clarity regarding how AI can create competitive advantage, while Zavodna et al. (2024) similarly find that firms in Central Europe struggle with low levels of knowledge, weak strategic vision, and uncertainty over expected

returns. The Serbian results, therefore, reinforce a recurring pattern observed in comparable economies. AI adoption is constrained less by the absence of technological capacity and more by deficits in managerial awareness, organizational culture, and the strategic framing of AI within business models.

Table 3. Reasons why companies do not use AI in business

Reason	Total (N=354)	Manufacturing (N=124)	Services (N=184)
AI not useful/relevant for business	237 (66.9%)	76 (61.3%)	131 (71.2%)
Legal uncertainty	38 (10.7%)	14 (11.3%)	16 (8.7%)
Lack of knowledge	34 (9.6%)	13 (10.5%)	14 (7.6%)
Concerns about data protection & privacy	28 (7.9%)	13 (10.5%)	10 (5.4%)
High costs	31 (8.8%)	12 (9.7%)	15 (8.2%)
Incompatibility with existing equipment/ software	27 (7.6%)	15 (12.1%)	9 (4.9%)
Difficulties with data availability/quality	27 (7.6%)	13 (10.5%)	8 (4.3%)
Don't know	23 (6.5%)	9 (7.3%)	12 (6.5%)
Ethical concerns	18 (5.1%)	12 (9.7%)	5 (2.7%)

Source: Own research

Regarding areas of use, the most common applications are data analysis and interpretation (41.4%), customer support (41.4%), and personalization of offers and marketing (38.6%). The services sector is particularly applying AI to customer support (48.7%), whereas manufacturing mainly uses it in personalization (44.6%) and data analysis and interpretation (41.1%). Other applications, such as automation of internal business processes (26.9%) and demand or customer behaviour prediction (23.4%), represent emerging areas of adoption, but remain below 30%. The least common applications are problem and risk detection (19.3%) and product development (21.4%), despite their significant potential for enhancing competitiveness (Table 4).

These sectoral patterns are consistent with international evidence showing that SMEs tend to adopt AI incrementally, concentrating first on applications that are less complex and offer visible short-term benefits. Customer-facing and data-analytic functions, such as marketing personalization and support services, are typically prioritized because they require fewer resources and generate rapid returns, whereas more transformative applications are often postponed (Proietti & Magnani, 2025; Sanchez et al., 2025). This gradualist approach reflects the cautious strategies of smaller firms operating under resource constraints, which favor low-risk, immediately applicable tools over long-term innovation projects (Schwaeke et al., 2025). In this sense, the Serbian results mirror broader tendencies observed in other SME contexts, yet the lower overall adoption rates compared to EU member states underline the additional structural and institutional limitations faced by enterprises in transition economies.

The results indicate that AI adoption in Serbia is still in an early stage, with sectors recognizing practical benefits but with a large share of enterprises perceiving AI as irrelevant to their operations. Future progress will depend on reducing informational barriers, strengthening employee capacities, and demonstrating concrete benefits through pilot projects and sector-specific use

cases. While the manufacturing sector shows slightly higher technological readiness, it also faces specific integration challenges with existing production systems. These findings contribute to the broader debate on digital transformation by providing evidence from a transition economy, with implications for both policymakers and business leaders.

Table 4. Areas of AI use among companies

Area of AI use	Total (N=145)	Manufacturing (N=56)	Services (N=76)
Data analysis & interpretation	60 (41.4%)	23 (41.1%)	31 (40.8%)
Customer support	60 (41.4%)	20 (35.7%)	37 (48.7%)
Personalization & marketing	56 (38.6%)	25 (44.6%)	26 (34.2%)
Automation of internal processes	39 (26.9%)	16 (28.6%)	18 (23.7%)
Forecasting demand / customer behaviour	34 (23.4%)	17 (30.4%)	16 (21.1%)
New product & service development	31 (21.4%)	18 (32.1%)	11 (14.5%)
Problem / error / risk detection	28 (19.3%)	13 (23.2%)	11 (14.5%)

Source: Own research

5. FUTURE RESEARCH DIRECTIONS

The findings of this study provide an important snapshot of the current state of AI adoption in Serbian enterprises, but they also raise new questions and highlight avenues for further inquiry. Several future research directions can be identified, both from an academic and a policy perspective.

While this study distinguishes between manufacturing and service enterprises, more granular analysis of subsectors is needed to uncover nuanced patterns. For example, within manufacturing, AI adoption in high-tech industries such as electronics or pharmaceuticals may differ significantly from traditional sectors such as food processing or textiles. Similarly, in services, the dynamics of AI adoption in financial services, retail, or logistics may vary depending on competitive pressures, customer demands, and regulatory environments. Future research should therefore explore sector-specific case studies to identify best practices and tailored support mechanisms.

The present study offers a cross-sectional view of AI adoption in 2025. Yet, AI is a fast-evolving technology, and adoption patterns are likely to shift significantly in the coming years as costs decrease, awareness grows, and government initiatives take effect. To fully capture these dynamics, future research should move beyond one-time assessments and instead establish mechanisms for continuous monitoring of firms over time. Such monitoring would make it possible to track how enterprises progress from initial experimentation to more advanced applications, what factors support sustained integration of AI, and under what circumstances firms reduce or abandon AI use. From a policy perspective, this type of evidence is particularly valuable. It would enable policymakers to evaluate the effectiveness of Serbia's national AI strategies (2020–2025; 2025–2030), measure the impact of infrastructure investments, and adjust support measures in response to observed bottlenecks. Systematic tracking of adoption trajectories could also inform the design of targeted interventions, such as awareness-raising campaigns, training programs, and sector-specific pilot projects, ensuring that public policies are responsive to the evolving needs

of enterprises. Embedding this kind of evidence-based monitoring into Serbia's broader digital transformation and innovation strategies would therefore not only enhance the relevance of public initiatives but also help align business adoption trends with national development objectives.

The findings indicate that many of the barriers to AI adoption are perceptual rather than purely technical, as a large share of firms consider AI irrelevant or not useful for their business models. This highlights the importance of organizational and behavioural factors, including managerial attitudes, leadership awareness, and the broader organizational culture, in shaping adoption outcomes. A deeper understanding of these dimensions can be gained by exploring how decision-makers interpret the role of AI in their strategic priorities and how employees perceive its implications for daily work practices. Approaches such as in-depth interviews and focus groups are particularly well suited to reveal these cognitive and cultural dynamics, offering insights that cannot be captured through survey data alone.

The skills gap has emerged as a recurrent theme in the international literature on AI adoption, often highlighted as a major obstacle to effective integration. While the present survey did not identify skills shortages as the most prominent barrier, the finding that nearly one in ten Serbian enterprises reported a lack of knowledge indicates that limited expertise remains a relevant issue. This suggests that workforce preparedness will be critical as AI technologies become more widespread. Understanding how enterprises build internal capacities, what types of training programs prove most effective, and how public policies can support upskilling and reskilling is therefore an important avenue for future investigation. In particular, examining the role of partnerships between firms, universities, and training providers could shed light on strategies for developing AI-related competencies in the Serbian context.

Positioning Serbia within a broader regional and international context is essential for understanding the dynamics of AI adoption. Comparative analyses with other Western Balkan economies can reveal whether Serbia is advancing more rapidly or facing similar challenges as its neighbours, while benchmarking against EU member states can highlight areas of convergence and divergence in adoption patterns. Such cross-country perspectives are valuable for identifying the influence of institutional frameworks, regulatory environments, and innovation ecosystems on technology uptake. At the same time, moving beyond descriptive accounts is crucial to capture the mechanisms through which enterprises perceive, adopt, and integrate AI technologies. Mixed-method approaches that combine survey-based evidence with in-depth case studies can provide richer insights into the interplay of technological, organizational, and institutional factors. In the Serbian context, this type of research would also support the evaluation of national AI strategies, inform the design of targeted sectoral initiatives, and contribute to aligning enterprise-level adoption with broader policy goals of digital transformation and economic modernization.

6. CONCLUSION

This paper has examined the adoption of AI among Serbian enterprises, focusing on adoption levels, barriers, and sectoral applications. Drawing on survey evidence from 499 firms, it provides one of the first systematic analyses of AI uptake in a transition economy.

The study shows that AI adoption in Serbia remains in an early stage, shaped less by technological readiness than by perceptions of relevance, organizational awareness, and managerial capacities. Firms that do adopt AI tend to concentrate on customer-facing and data-analytic applications, while more transformative uses remain limited. These patterns are consistent with findings from

other SME contexts, but adoption levels in Serbia are comparatively lower, reflecting the structural and institutional constraints of a transition economy.

The analysis carries important policy implications. Overcoming informational barriers, strengthening skills, and demonstrating concrete business value will be as critical as investments in infrastructure or financial incentives. Sectoral differences further suggest that tailored support measures are needed to reflect the distinct challenges of manufacturing and services.

By situating Serbia within the broader debate on digital transformation, this study highlights both the universal challenges of AI adoption and the specific obstacles faced by firms in smaller and emerging markets. Its findings are relevant not only for national policymakers but also for regional actors across the Western Balkans, where similar barriers persist. The effective realization of Serbia's AI strategies will depend on aligning policy frameworks with enterprise-level needs, fostering an environment in which AI can serve as a driver of competitiveness, innovation, and sustainable growth.

Acknowledgment

The research presented in this paper was funded by the Ministry of Science, Technological Development and Innovation of the Republic of Serbia under contract number 451-03-136/2025-03.

References

Cannavale, C., Claudio, L., & Koroleva, D. (2025). Digitalisation and artificial intelligence development: A cross-country analysis. *European Journal of Innovation Management*, *28*(11), 112-130. https://doi.org/10.1108/EJIM-07-2024-0828

Chen, Y., & Jin, S. (2023). Artificial intelligence and carbon emissions in manufacturing firms: the moderating role of green innovation. *Processes*, *11*(9), 2705. https://doi.org/10.3390/pr11092705

Czarnitzki, D., Fernandez, G. P., & Rammer, C. (2023). Artificial intelligence and firm-level productivity. *Journal of Economic Behavior & Organization*, *211*, 188–205. https://doi.org/10.1016/j.jebo.2023.05.008

Enshassi, M., Nathan, R., Soekmawati, S., Al-Mulali, U., & Ismail, H. (2024). Potentials of artificial intelligence in digital marketing and financial technology for small and medium enterprises. *IAES International Journal of Artificial Intelligence (IJ-AI)*, *13*(1), 639-647. https://doi.org/10.11591/ijai.v13.i1.pp639-647

European Commission. (2020). White Paper on Artificial Intelligence - A European approach to excellence and trust. https://commission.europa.eu/system/files/2020-02/commission-white-paper-artificial-intelligence-feb2020_en.pdf

Ghani, E., Ariffin, N., & Sukmadilaga, C. (2022). Factors influencing artificial intelligence adoption in publicly listed manufacturing companies: a technology, organisation, and environment approach. *International Journal of Applied Economics Finance and Accounting*, *14*(2), 108-117. https://doi.org/10.33094/ijaefa.v14i2.667

Gładysz, B., Matteri, D., Ejsmont, K., Corti, D., Bettoni, A., & Haber, R. (2023). Platform-based support for ai uptake by SMEs: guidelines to design service bundles. *Central European Management Journal*, *31*(4), 463-478. https://doi.org/10.1108/cemj-08-2022-0096

Kassa, B. Y., & Worku, E. K. (2025). The impact of artificial intelligence on organizational performance: The mediating role of employee productivity. *Journal of Open Innovation: Technology, Market, and Complexity*, *11*(1), 100474. https://doi.org/10.1016/j.joitmc.2025.100474

Kuzminov, M. (2024). Strategic management and artificial intelligence as tools for effective activity of enterprises. *Economies' Horizons*. 145-152. DOI: 10.31499/2616-5236.4(29).2024.316528.

Li, C., Xu, Y., Zheng, H., Wang, Z., & Li, C. (2023). Artificial intelligence, resource reallocation, and corporate innovation efficiency: Evidence from China's listed companies. *Resources Policy*, *81*, 103045. https://doi.org/10.1016/j.resourpol.2023.103324

Nesterova, K. (2024). Integration of artificial intelligence into strategic business management: development prospects. *Herald UNU International Economic Relations and World Economy*. 10.32782/2413-9971/2024-53-13.

OECD. (2025). Governing with Artificial Intelligence: The State of Play and Way Forward in Core Government Functions, OECD Publishing, Paris, https://doi.org/10.1787/795de142-en.

OECD/BCG/INSEAD. (2025). The Adoption of Artificial Intelligence in Firms: New Evidence for Policymaking, OECD Publishing, Paris. https://doi.org/10.1787/f9ef33c3-en

Oyekunle, D., & Boohene, D. (2024). Digital transformation potential: the role of artificial intelligence in business. *International Journal of Professional Business Review*, *9*(3), e04499. https://doi.org/10.26668/businessreview/2024.v9i3.4499

Pariso, P., Picariello, M., & Marino, A. (2025). How AI can aim and reduce the annual amount of energy consumption in European SMEs. *Macromolecular Symposia*, *414*(3) https://doi.org/10.1002/masy.70045

Proietti, S., & Magnani, R. (2025). Assessing AI Adoption and Digitalization in SMEs: A Framework for Implementation. *arXiv preprint*. https://arxiv.org/abs/2501.08184.

Sanchez, E., Calderon, R., & Herrera, F. (2025). Artificial Intelligence Adoption in SMEs: Survey Based on TOE–DOI Framework, Primary Methodology and Challenges. *Applied Sciences*, *15*(12), https://doi.org/10.3390/app15126465

Schwaeke, J., Peters, A., Kanbach, D. K., Kraus, S., & Jones, P. (2025). The new normal: The status quo of AI adoption in SMEs. *Journal of Small Business Management, 63*(3), 1297-1331. https://doi.org/10.1080/00472778.2024.2379999

Tornatzky, L. G., & Fleischer, M. (1990). The processes of technological innovation. Lexington, MA: Lexington Books.

Wang, Y., & Liu, F. (2025). Impact of artificial intelligence innovation on food company performance. *International Review of Financial Analysis*, 103, 104219. https://doi.org/10.1016/j.irfa.2025.104219

Wıjayatı, D., Rahman, Z., Fahrullah, A., Rahman, M., Arifah, I., & Kautsar, A. (2022). A study of artificial intelligence on employee performance and work engagement: the moderating role of change leadership. *International Journal of Manpower*, 43(2), 486-512. https://doi.org/10.1108/ijm-07-2021-0423

World Economic Forum. (2025). AI in Action: Beyond Experimentation to Transform Industry. https://reports.weforum.org/docs/WEF_AI_in_Action_Beyond_Experimentation_to_Transform_Industry_2025.pdf

Xue, M., Cao, X., Xu, F., Gu, B., & Zhang, Y. (2022). The effect of artificial intelligence (AI) on firm labor structure. Proceedings of the 55th Hawaii International Conference on System Sciences, https://doi.org/10.24251/hicss.2022.853

Zavodna, L. S., Uberwimmer, M., & Frankus, E. (2024). Barriers to the implementation of artificial intelligence in small and medium-sized enterprises: Pilot study. *Journal of Economics and Management*, 46(1), 331-352. https://doi.org/10.22367/jem.2024.46.13

Appendix A. Survey questions on artificial intelligence

Q1. In which areas does your company use artificial intelligence (AI)?
(Multiple answers allowed)

- ☐ Data analysis and interpretation (e.g. use of AI for pattern recognition, data-driven decision-making)
- ☐ Personalisation of offers and marketing activities (e.g. product/service recommendations or targeted marketing campaigns based on customer behaviour)
- ☐ Customer support (e.g. chatbots, automated responses to customer inquiries, virtual assistants)
- ☐ Forecasting demand and/or customer behaviour (e.g. sales forecasting, analysis of customer behaviour or user needs)
- ☐ Automation of internal business processes (e.g. automation of reporting, document processing, procurement processes)
- ☐ Detection of problems, errors, or risks (e.g. AI systems for fraud detection, fault detection, or identification of operational inefficiencies)
- ☐ Development of new products and services (e.g. use of AI for idea generation, prototype testing, or design optimisation)
- ☐ We do not use artificial intelligence
- ☐ Other (please specify): _____

Q2. If your company does not use artificial intelligence, what are the main reasons?
(Multiple answers allowed)

- ☐ High costs
- ☐ Lack of knowledge or expertise
- ☐ Incompatibility with existing equipment, software, or systems
- ☐ Difficulties related to data availability or data quality
- ☐ Concerns about data protection and privacy
- ☐ Uncertainty regarding legal implications
- ☐ Ethical concerns related to the use of AI technologies
- ☐ AI technologies are not useful or relevant for the company
- ☐ Other (please specify): _____
- ☐ Do not know

Balkans JETSS (2025) 2: 80-89
DOI: https://doi.org/10.31410/Balkans.JETSS.2025.8.2.80-89

Original Scientific Article

Should I Stay or Should I Go? Dilemmas and Polarized Attitudes of Public Sector Managers Regarding AI

Irina Teodora Manolescu[1]
Dan Vîrlan[2]
Mihai Talmaciu[3]

Received: October 10, 2025 / Revised: December 29, 2025 /
Accepted: December 30, 2025 / Published: December 30, 2025

Abstract: *The public sector, characterized by structural inertia, strict regulations, and a low level of flexibility, faces additional challenges in relation to the major technological transformations of recent years, particularly those associated with artificial intelligence (AI). The research aims to investigate the perceptions of managers in the Romanian public sector regarding the opportunities and risks brought by the implementation of AI in organizations. The study employs a mixed-method design, combining two complementary approaches: a survey of 188 managers from public institutions and a focus group involving 7 managers from various public administration organizations. The results reveal a set of recurring trends. In the absence of a clear institutional vision, AI is perceived more as a source of risk than as a strategic opportunity. The study makes an important contribution to understanding this complex landscape by highlighting both the barriers and the opportunities perceived by public administration managers in relation to AI.*

Keywords: *Artificial Intelligence, Public management, Technological transformation*

JEL Classification: H11 · O33

✉ irina.manolescu@gmail.com
[1] "Alexandru Ioan Cuza" University of Iasi, Faculty of Economics and Business Administration, Carol I Bd., 22, 700505, Iasi, Romania
[2] "Alexandru Ioan Cuza" University of Iasi, Faculty of Economics and Business Administration, Carol I Bd., 22, 700505, Iasi, Romania
[3] "Alexandru Ioan Cuza" University of Iasi, Faculty of Economics and Business Administration, Carol I Bd., 22, 700505, Iasi, Romania

1. INTRODUCTION

Artificial intelligence (AI) is no longer perceived as a static collection of rigid algorithms, but as a dynamic system capable of learning and adapting, redefining the way organizational processes - including those within public management - can be designed and optimized in the digital era. AI represents one of the most spectacular technological transformations of recent decades, with major implications for how organizations operate, make decisions, and adapt their internal processes. However, the magnitude and direction of these changes are difficult to anticipate, and their concrete effects depend both on the organizational context and on the capacity of institutional actors to integrate new technologies. The public sector, characterized by structural inertia, strict regulations, budgetary constraints, and a low level of flexibility, faces additional challenges in relation to these transformations. At the governmental level, responses regarding the use of AI in administration have varied radically - from temporary bans on AI use in public institutions to initiatives that actively encourage accelerated adoption, such as subsidizing AI subscriptions for the entire population. Similarly, at the level of managers and employees within public institutions, perceptions are highly fragmented. A deeper understanding and critical analysis of how managers form their perceptions of AI-based technologies represent important prerequisites for ensuring transparent, sustainable, and ethically aligned adoption within management processes. Managerial perception influences not only the decision to implement AI but also the degree to which it is accepted and integrated institutionally. Trust in the benefits and accuracy of AI systems can act as a catalyst for effective institutional integration; by contrast, skepticism or distrust may slow down or even block the implementation of these advanced technologies. AI has the potential to increase efficiency, reduce costs, and optimize processes; it can also support managers in data-driven decision-making and contribute to greater transparency and compliance with internal policies (Chukwuka & Dibie, 2024). The main concern, since the large-scale operationalization of AI tools, remains the potential reduction - at least partially - of certain jobs. In this context, an important aspect to consider is employees' level of awareness regarding human–machine collaboration (Nawaz et al., 2024; Onea Neculăesei & Manolescu, 2024; Pratiwi, 2024). Managers at all organizational levels will increasingly need to adapt to the world of intelligent machines in order to maintain competitiveness (Yawar & Hakimi, 2025). The major challenge remains achieving a functional and ethical balance between the two dimensions (Pelea, 2019). This may lead to a complete reconfiguration of human resource management strategies, with implications for employee reskilling, digital competence development, and the formulation of new institutional policies adapted to the AI era. The automation of HRM decisions through AI can raise issues of fairness and transparency, as decisions such as recruitment, promotion, and evaluation may be influenced by algorithms that are opaque to employees and community members, increasing suspicions of manipulation or discrimination - issues already present in the public sector. The specialized literature highlights potential risks associated with AI use, one of the most significant being its potential to perpetuate or even amplify existing biases (Chatterjee et al., 2024). Despite ongoing initiatives, AI implementation remains insufficiently regulated. Thus, the foundations for mechanisms that ensure the ethical application of AI are still lacking (Stuss & Fularski, 2024).

The relevance of this research topic stems from the following rationale: without examining perceptions at the grassroots level, a disruptive technology is unlikely to be effectively adopted. Consequently, there is a risk of bypassing essential developmental stages. For instance, in certain municipalities, the lack of attention to employees' resistance to change has resulted in computers still being used merely as typewriters, with digitalization remaining largely unfamiliar. Within such a context, attempts are being made to introduce AI tools even though the prerequisite stages of technological adaptation have not yet been completed. The failure to acknowledge and address resistance to change thus emerges as a critical factor contributing to the unsuccessful implementation of major transformation initiatives.

2. THE CONTEXT OF AI IN THE ROMANIAN PUBLIC ADMINISTRATION

The general context begins with the opinions of renowned figures in the field of technology, many of whom expressed strong negative views as soon as AI became a quasi-universal paradigm. Although opinions have been divided, this underlying negativity has the potential to generate resistance at both organizational and individual levels. Negative narratives about AI have spread across multiple domains - unlike previous innovations, whose effects tended to remain more contained within specific sectors. The implementation of AI in public management is a relatively new and little-known trend in Romania. Consequently, research on this topic remains scarce - a search in the Web of Science database reveals only 16 articles focused on AI in the Romanian public sector. As with any disruptive change, AI brings both opportunities and challenges, requiring careful analysis of managers' perceptions regarding trust in AI systems and the adaptability of institutions to technological advancements. Managerial perceptions of technology vary depending on experience and accumulated knowledge, the institution's infrastructure, and the existing regulatory framework - the latter playing a key role in ensuring an atmosphere of trust and security within managerial processes (Maragno et al., 2023). Context is therefore crucial in shaping AI implementation plans. The benefits (efficiency, cost reduction, decision-making objectivity) and the challenges and risks (poor infrastructure, lack of digital skills, resistance to change, algorithmic bias) form the main pillars underpinning arguments for proposed transformations. The specific context of the Romanian public sector with respect to AI implementation is not a favorable one. The Digital Economy and Society Index (DESI) indicators consistently place Romania at the bottom of the EU ranking, at a considerable distance from the European average (Figure 1). Although a national AI strategy exists - with a dedicated chapter for the public sector (Autoritatea Națională pentru Cercetare, 2024) - subsequent sector-specific regulations are still missing.

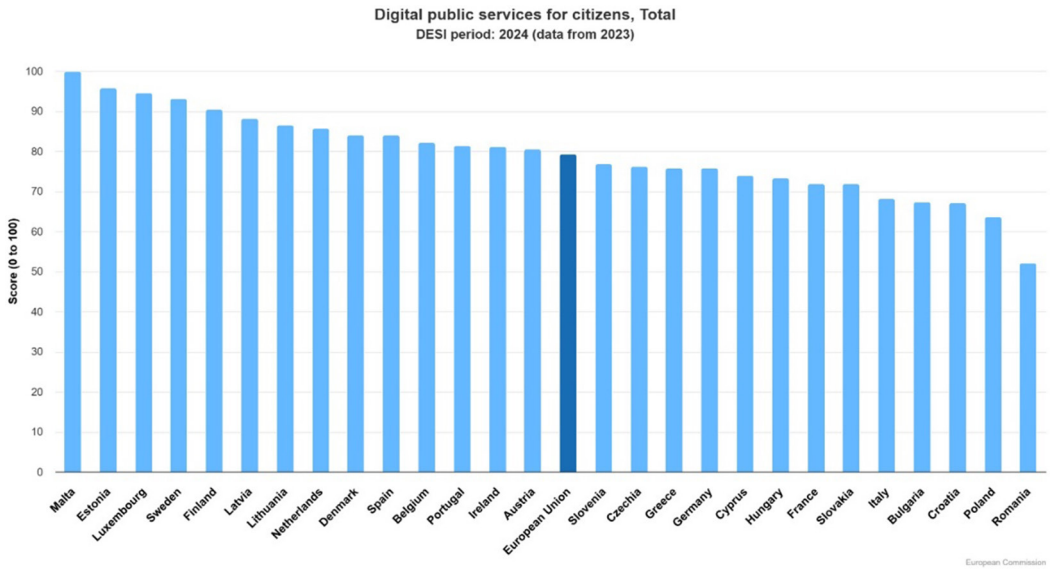

Figure 1. The hierarchy of EU countries by the relevant DESI indicator
Source: European Commission (2025)

A highly publicized initiative that contributed to widespread distrust in the effectiveness of AI at the national level was the introduction, for the first time in early 2023, of an AI-based government adviser named *ION*, designed to enhance transparency and communication with citizens

(Autoritatea Națională pentru Cercetare, 2023). The initiative, however, was insufficiently prepared and poorly adapted to the specific administrative processes. After a year marked by numerous controversies, the AI adviser was ultimately "dismissed." The impact of this initiative was significant - it generated public distrust on the one hand, but also helped to mitigate fears regarding the potential replacement of human resources by AI tools. Nonetheless, there are examples of good practices in other countries that could be considered and adapted to the Romanian context. The main directions for using AI tools in the public sector, as outlined by Van Noordt and Misuraca (2022), include policy development, service delivery, and internal management. AI is used to detect social problems, simulate policy scenarios, and evaluate their implementation. In public service delivery, the most common applications are chatbots and virtual assistants, which help personalize services, automate repetitive tasks, and create proactive service models. In internal management, AI is employed for human resource allocation, fraud and corruption detection, cybersecurity, and public procurement. Collaboration with the private sector is essential for successful AI adoption in the public sphere, as the synergy between the two paradigms enhances the feasibility and effectiveness of results (Talmaciu et al., 2023).

The factors that can stimulate AI adoption in Romania are technological (increased access to open-source and cloud-based AI solutions, accelerated digitalization of public services), organizational (pressure for efficiency, reduced staff), and environmental (European requirements for digitalization and interoperability, availability of EU funds). However, these same factors can also hinder progress (Alamäki, 2025) - with the lack of digital skills and collaborative environments being among the most critical obstacles. National strategies related to AI tend to emphasize efficiency and service delivery (Hjaltalin & Sigurdarson, 2024) while paying less attention to citizen engagement and democratic values. Prioritization remains the key approach to improving the efficiency of the public sector under significant budgetary constraints (Manolescu et al., 2024), and AI implementation stands as an unquestionable global priority.

3. METHODOLOGY

This research aims to investigate the attitudes and perceptions of managers in the public sector regarding the opportunities and risks brought by the implementation of AI in organizations. The study focuses on several key questions concerning the integration of AI into managerial processes. The investigation was structured around several specific research goals: (1) Identifying managers' perceptions regarding the potential benefits of AI implementation in Romanian public institutions; (2) Identifying perceptions of the challenges and risks associated with AI implementation; (3) Assessing the level of preparedness of Romanian public institutions for integrating AI-based technologies into management processes, from the perspective of managerial perception.

The study employed a mixed-method research design, combining two complementary approaches: a survey administered to a broad sample of 188 public-sector managers and a focus group involving 7 managers from diverse public administration organizations. The purpose of using this methodological mix was to capture not only the level of preparedness and general attitudes but also the nuances of internal debates concerning the implementation of this technology, including concerns about potential workforce reductions due to automation. The main data collection instrument was a structured questionnaire designed to capture both relevant demographic information and existing or potential opinions, attitudes, and practices related to AI use in the public sector. The questionnaire consisted of 24 items, varying in format to ensure comprehensive data collection: closed-ended questions, multiple-response questions, Likert-scale items (1 to 5) and an open-ended question. The target population consisted of managers from Romanian public

institutions, ensuring the relevance of the collected opinions for the study's objectives, given their decision-making role in technological adoption processes. Data collection was conducted online via Google Forms, with the questionnaire distributed to over 10,000 email addresses belonging to various types of public institutions across Romania. As a result of this process, conducted between January and June 2025, a total of 188 valid responses were obtained and used for statistical analysis. While the response rate was relatively low (below 2%), which may represent a limitation, the heterogeneity of the sample and its final size nonetheless support the robustness of the findings. The sample structure was diverse in terms of institutional representation - most respondents were from municipalities (56.5%) and educational institutions (22%), while others represented county councils, deconcentrated agencies, central institutions, healthcare units, and national or international missions. Regarding institutional environment, the distribution was relatively balanced, with a slight predominance of respondents from rural areas (53.7%). Most organizations had fewer than 50 employees (52.2%). The respondents' age and managerial experience were also evenly distributed, with the majority falling within the 46–55 age group (37.8%) and having 6–15 years of managerial experience (33.9%). The focus group was conducted in July 2025, with the voluntary participation of seven managers from different public institutions (Table 1), all attending a professional development program. The sample size was deliberately set to ensure a range of organizational perspectives while maintaining adequate time for individualized intervention. Two discussion sessions were held over one week - the first focused on framing and problem-setting, and the second on debate and reflection. The key discussion themes mirrored those addressed in the survey (level of knowledge and use, benefits and risks), while also providing deeper insights into organizational and individual approaches.

Table 1. Characteristics of Focus Group Participants

Type of organization	Position	Gender	Managerial experience (years)
Rural Municipality	Department Manager	M	5
Urban Municipality	Service Manager	F	22
County Council	Department Manager	M	14
Regional Development Agency	Department Manager	M	7
Water Management System	Office Manager	M	8
County Administration of Public Finances	Department Manager	F	12
County Environmental Guard (National Environmental Agency)	Team Manager	M	2

Source: Own processing

The data processing and analysis integrated both quantitative and qualitative approaches. The results were then compared with those of other studies addressing complex issues related to the use of AI in the public sector. The analyses were limited to descriptive statistics (means, skewness, and kurtosis). Inferential analyses, such as correlation analyses, were not conducted, as the primary aim of the study was exploratory and focused on characterizing the data distribution.

4. RESULTS AND DISCUSSION

The data analysis provides a complex and nuanced picture of respondents' perceptions and attitudes toward AI, while also revealing a series of recurring trends (Table 2). The key findings are as follows: almost all respondents (83.5%) identified at least one benefit of AI in recruitment; more than half of the surveyed managers reported having already used generative AI tools; the average level of knowledge about AI was approximately 3.05 on a scale from 1 to 5, indicating a

general familiarity with the concept; nearly half of the interviewed managers (48.4%) expressed willingness to implement AI but emphasized a lack of adequate resources or competencies; support for data-driven decision-making had the most significant impact among the perceived benefits related to employee motivation and job satisfaction; a well-defined regulatory and ethical framework enhances the willingness to adopt AI; managers who view AI as a complementary tool supporting specialists rather than replacing them are considerably more open to organizational change; approximately 70% of respondents expressed skepticism or distrust toward the idea of a centralized national AI-based system - this indicates a preference for gradual and decentralized implementations, with direct institutional involvement; over half of respondents (56.9%) reported obtaining information about AI primarily from informal sources (such as social media and popular news outlets) rather than formal or scientific ones. Managers informed through validated, science-based sources exhibited significantly more positive and responsible attitudes toward AI implementation.

Table 2. Summary of the descriptive analyses of the core variables of the study

Analyzed variable	Frequency / Percentage analysis	Descriptive indicators	Interpretation
Level of Knowledge about AI	Mean score is 3.05 (on a 1–5 scale)	Mean = 3.05	Respondents have a medium to good level of familiarity with the concept of artificial intelligence.
Sources of Information about AI	56.9% Informal sources, 43.1% Formal sources	Mean = 1.57, Skewness = -0.282	A slight preference for informal sources of information is observed, indicating a need for professionalization and access to academic or specialized sources.
Use of Generative AI	52.1% have used it, 10.1% intend to use it, 37.2% have not used it	Mean = 1.59, Skewness = 0.847	High degree of adoption and technological openness. The positive skewness shows that although many have used it, there is still a significant segment that has not.
Perception of AI Benefits	83.5% identified at least one benefit (multiple responses, with each option transformed into a distinct binary variable (0 = no, 1 = yes).	Mean = 1.16, Skewness = 1.821	Very high receptiveness. The vast majority of respondents recognize AI's advantages, and the strongly positive skewness shows that very few saw no benefit at all.
Perception of AI Risks	Balanced distribution: Low Risk (28.7%), Medium (27.7%), High (27.1%)	Mean = 2.31, Skewness = 0, Kurtosis = -1.204	Opinions regarding risks are polarized and dispersed, with no dominant trend. Negative kurtosis confirms the lack of concentration around a single opinion.
Opinion on a National AI System	69.7% are skeptical or disagree, 30.3% are confident	Mean = 1.70, Skewness = -0.863	The majority tendency is skepticism towards a centralized solution, despite the general openness to technology.

Source: Own calculations

The majority of respondents are key decision-makers in institutions with ongoing digitalization initiatives and recognize the advantages of AI, which indicates a solid potential for adoption. However, despite the clear willingness to implement AI, barriers related to resources and competencies continue to hinder progress. Moreover, opinions regarding AI-related risks are polarized, and skepticism toward a centralized national AI system remains dominant, suggesting a preference for decentralized solutions or a greater need for trust and clarity in national-level implementation. Managers emphasized the lack of necessary resources for effective implementation - from technological infrastructure and financial support to dedicated training programs. This discrepancy is also reflected in the literature; for example, in India, while 89% of public organizations had ongoing AI projects, only 25% possessed advanced technological infrastructure (Kulal et al.,

2024). In the absence of a clear institutional vision, AI tends to be perceived more as a source of risk than as a strategic opportunity. Another phenomenon observed is technological determinism - the implementation of AI driven by technological availability rather than by actual institutional needs (Mergel et al., 2024).

The qualitative research also identified specific areas in which managers perceive immediate advantages. These include the automation of repetitive processes, improvement of service quality for citizens, and enhanced internal efficiency through reduced administrative workloads. Such benefits could serve as starting points for a gradual implementation strategy, provided that obstacles related to professional preparedness and resource allocation are addressed as priorities. The analysis shows that although public managers are individually adopting AI to strengthen citizen trust, large-scale implementation is hampered by skepticism toward centralized systems, with success depending critically on formal education and transparent trust-building strategies. The results reflect a major concern regarding the impact of AI-based technologies on citizen relations, administrative control, and data protection, while underscoring the importance of clear, transparent, and balanced public policies for AI adoption in the public sector. The findings offer a clear perspective on general trends and a growing consensus around the idea of responsible AI use. The results of this research outline a nuanced picture of Romanian public managers' perceptions of AI implementation, highlighting both alignment with global trends described in the literature and specific local characteristics. Broadly, the findings support existing studies suggesting that efficiency-related immediate benefits are the main driver of AI adoption in organizational management. The managers surveyed expressed particular enthusiasm for automating repetitive tasks and accelerating recruitment processes - findings consistent with numerous studies emphasizing AI's potential to handle large volumes of administrative work and free up specialists' time for strategic or creative activities. The fact that long-term cost reduction and improved accuracy in candidate selection did not appear as strongly influential in managers' perceptions may seem surprising at first, since the literature often emphasizes these benefits (Talmaciu & Manolescu, 2023). A possible explanation lies in the time horizon and tangibility of benefits - managers tend to be more responsive to immediate, concrete advantages than to diffuse or future gains. This pragmatic orientation is also evident in their approach to risk: most managers acknowledge a moderate number of risks, adopting a realistic but not alarmist stance toward technology adoption. They anticipate potential challenges but do not appear fundamentally discouraged by them, reflecting an openness to innovation tempered by rational caution. A key finding of the study is the perceived importance of AI-supported decision-making, with managers valuing AI's ability to provide better information for decisions. This aligns with a growing body of literature viewing AI in organizations as a form of "augmented reality" (Onea Neculăesei & Manolescu, 2024) - an instrument that complements and enhances human decisions rather than replacing them. The fact that AI benefits related to managerial decision-making had the strongest impact on openness suggests that public managers are willing to become "augmented managers", using AI-based tools to increase the accuracy and effectiveness of their decisions. Conversely, AI benefits oriented toward employees (such as personalized training or objective evaluation) were less convincing to managers. This may indicate that, at the current stage, managerial culture remains more focused on structural efficiency than on employee development. As public managers are accountable for organizational performance as a whole, they may perceive AI adoption primarily in relation to personal satisfaction and motivation rather than workforce empowerment. Additionally, in the public sector, certain constraints may shape how AI influences employee experience.

In the Romanian context, the data indicate that public policies focused solely on technological acquisition or superficial digitalization will have limited impact unless accompanied by an innovation-oriented culture and human capital development within institutions. This underscores the

need for integrated organizational change strategies (Bulat et al., 2024). One of the most interesting and relevant findings concerns the ethical and regulatory framework. Managers conveyed a clear message of willingness to implement AI - but in a responsible and secure manner. Consequently, developing explicit institutional ethical codes could accelerate AI adoption. The focus on sustainability and avoidance of unprepared initiatives was another strong theme, consistent with prior research (Manolescu & Talmaciu, 2021; Wilson & Van Der Velden, 2022). Five essential boundary conditions for social sustainability in AI were identified: diversity, learning capacity, self-organization capacity, common meaning and trust.

Findings on resistance to change suggest that the narrative of AI as a threat is losing ground among elected public leaders, being replaced by a more positive, collaborative outlook. Managers who perceive AI as a tool that complements their teams and believe in human–machine co-creation are far more likely to implement AI successfully. While many studies describe AI as a threat to the workforce due to its increasingly advanced capabilities, paradoxically, these same capabilities are also perceived as benefits (Ban et al., 2024). The fact that fear or potential losses no longer dominate managerial discourse - replaced instead by narratives of collaboration and opportunity - signals a maturing perspective. AI can upskill employees and improve decision-making processes, but may also generate resistance to change and trust deficits (Maragno et al., 2023). However, recent global research shows that employees still experience high levels of anxiety about AI: approximately 77% of workers are concerned about job loss due to AI (McGraw, 2024). In this context, public managers must navigate the pressures associated with the pace of institutional change. The Romanian results indicate that reluctance toward AI is no longer primarily driven by fears of job cuts or the pace of transformation, but rather by how change is communicated and managed. Institutional resilience and individual adaptability can be fostered through proactive communication and human resource development policies. The greatest concern expressed in the focus group was a realistic one: that recruitment and selection might place excessive emphasis on AI-related skills, leading to the loss of other complementary and essential competencies. Perceived risks resonate strongly with current academic and public debates. These concerns should be interpreted not as opposition to AI, but as areas requiring attention during implementation. Managers did not reject AI because of these risks but instead outlined a set of conditions they deemed essential. Similarly, the perceived risk of technological dependence underscores the importance of backup plans and redundancy, with managers implicitly signaling that their institutions are not yet ready to become fully digital overnight. Instead, capacity should be built gradually - supported by reliable infrastructure, manual backups for emergencies, and preservation of internal competencies.

5. CONCLUSION

In the public sector, perceptions of AI oscillate between a desire for modernization and a fear of the unknown. The lack of a coherent institutional and national direction amplifies uncertainty, and without concerted interventions - through clear public policies, investment, and professional training - there is a risk that this sector will lag behind others in adopting new technologies. Romanian public managers are generally prepared to adopt AI, but success depends on a responsible approach to implementation, the provision of necessary resources and competencies, the promotion of an innovative organizational culture, and the establishment of a clear ethical framework. The successful and sustainable integration of AI in the Romanian public sector requires a holistic, human-centered, ethically grounded, and strategically coordinated approach. This goes beyond mere technological adoption, involving an organizational and cultural transformation within each institution. Given the open yet cautious attitude of public managers toward artificial intelligence in human resource management, several practical and strategic recommendations are proposed for decision-makers in the

Romanian Government, the Authority for Digitalization of Romania, public administration bodies, and other relevant institutions: prioritize investments in resources and competencies; develop and communicate a robust ethical framework for AI; promote the tangible operational benefits of AI; foster an innovative and adaptable organizational culture; invest in reskilling and human–machine collaboration programs; exercise caution and ensure consultation in developing centralized national AI-based recruitment and selection systems; promote formal information sources about AI. These recommendations synthesize the overarching framework derived from the detailed analysis of Romanian public managers' perceptions. By taking these into account, public institutions can fully harness the transformative potential of artificial intelligence in public administration, ensuring a responsible, efficient, and citizen-centered digital transition.

References

Alamäki, A. (2025). Expanding AI adoption in public sector organizations: perspectives on management practices. *Transforming Government: People, Process and Policy*. https://doi.org/10.1108/TG-05-2025-0124

Autoritatea Națională pentru Cercetare. (2023). *Romanian premiere: ION, the first government advisor in the world to use artificial intelligence* [Press release]. https://www.research.gov.ro/premiera-romaneasca-ion-primul-consilier-guvernamental-din-lume-ce-va-folosi-inteligenta-artificiala-9546/

Autoritatea Națională pentru Cercetare. (2024). *National strategy in the field of artificial intelligence 2024 - 2027*. https://www.research.gov.ro/programe-nationale/strategia-nationala-in-domeniul-inteligentei-artificiale-2024-2027/

Ban, O., Maiorescu, I., Bucur, M., Sabou, G. C., & Cohen Tzedec, B. (2024). AI between Threat and Benefactor for the Competencies of the Human Working Force. *Amfiteatru Economic, 26*(67), 762-782. https://doi.org/10.24818/EA/2024/67/762

Bulat, C. I., Roman, C. T., & Manolescu, I. T. (2024). Insights into organizational change dynamics in higher education institutions. *Cross-Cultural Management Journal, 26*(1), 7-17. https://doi.org/10.70147/c26717

Chatterjee, S., Jemima, A. G., Ray, S., Kumar, S., & Ahluwalia, G. (2024). A systematic review of artificial intelligence (AI) and impact on human resource management (HRM): Challenges, risks, and opportunities. *Naturalista Campano*, *28*(1), 558-591.

Chukwuka, E. J., & Dibie, K. E. (2024). Strategic Role of Artificial Intelligence (AI) on Human Resource Management (HR) Employee Performance Evaluation Function. *International Journal of Entrepreneurship and Business Innovation, 7*(2), 269-282.

European Commission. (2025). DESI indicators. [Data visualization]. https://digital-decade-desi.digital-strategy.ec.europa.eu/datasets/desi/charts/desi-indicators?period=desi_2024&indicator=desi_dps_cit&breakdown=total&unit=egov_score&country=AT,BE,BG,HR,CY,CZ,DK,EE,EU,FI,FR,DE,EL,HU,IE,IT,LV,LT,LU,MT,NL,PL,PT,RO,SK,SI,ES,SE

Hjaltalin, I. T., & Sigurdarson, H. T. (2024). The strategic use of AI in the public sector: A public values analysis of national AI strategies. *Government Information Quarterly*, *41*(1), 101914. https://doi.org/10.1016/j.giq.2024.101914

Kulal, A., Rahiman, H. U., Suvarna, H., Abhishek, N., & Dinesh, S. (2024). Enhancing public service delivery efficiency: Exploring the impact of AI. *Journal of Open Innovation: Technology, Market, and Complexity*, *10*(3), 100329. https://doi.org/10.1016/j.joitmc.2024.100329

Manolescu, I. T., Medeleanu, C. M., & Talmaciu, M. (2024). The unseen consequences of prioritization. Dilemmas and investment security in European initiatives. In: N. Davydenko (Ed.) *Economic and financial sequrity of the state: international aspect* (pp. 368-378). Nuremberg: Verlag.

Manolescu, I. T., & Talmaciu, M. (2021). The Involvement of Local Action Groups in Regional Sustainable Development - A Multi-Stakeholder Analysis. In: M. Ivanova, D. Nikoloski, R. Yilmaz (Eds.) *Proceedings of XV. IBANESS Congress Series on Economics, Business and Management* (pp. 210-221).

Maragno, G., Tangi, L., Gastaldi, L., & Benedetti, M. (2023). Exploring the factors, affordances and constraints outlining the implementation of Artificial Intelligence in public sector organizations. *International Journal of Information Management, 73*, 102686. https://doi.org/10.1016/j.ijinfomgt.2023.102686

McGraw, M. (2024). Research Finds Understanding of GenAI Lags Among Public Sector Workforce. PSHRA. https://pshra.org/research-finds-understanding-of-genai-lags-among-public-sector-workforce/

Mergel, I., Dickinson, H., Stenvall, J., & Gasco, M. (2024). Implementing AI in the public sector. *Public Management Review*, 1-14. https://doi.org/10.1080/14719037.2023.2231950

Nawaz, N., Arunachalam, H., Pathi, B. K., & Gajenderan, V. (2024). The adoption of artificial intelligence in human resources management practices. *International Journal of Information Management Data Insights*, *4*(1), 100208. https://doi.org/10.1016/j.jjimei.2023.100208

Onea Neculăesei, A. N., & Manolescu, I. T. (2024). Humanistic Management and Human-Technology Interaction. *CrossCultural Management Journal*, *26*(2), 105-114. https://doi.org/10.70147/c26105114

Pelea, C. I. (2019). The relationship between artificial intelligence, human communication and ethics. A futuristic perspective: Utopia or dystopia? *Media Literacy and Academic Research, 2*(1), 38-48.

Pratiwi, Y. N. D. (2024). The Role of Science Artificial Intelligence for Trend of Digital HRM. *Jurnal Penelitian Pendidikan IPA*, *10*(12), 914-919. https://doi.org/10.29303/jppipa.v10i12.9421

Stuss, M. M., & Fularski, A. (2024). Ethical considerations of using artificial intelligence (AI) in recruitment processes. *Education of Economists and Managers, 71*(1), 53-67. https://doi.org/10.33119/EEIM.2024.71.4

Talmaciu, M., & Manolescu, I. T. (2023). The multi-stakeholder network as policy tool for a robust governance of the public organizations in turbulent times. In: A.M. Bercu, I. Bilan, C.M. Apostoaie (Eds.). *Elevating Europe. Smart Initiatives and Administrative Innovation. Proceedings of the International Conference EU-PAIR* (pp. 273-287), Iaşi: Universităţii „Alexandru Ioan Cuza" din Iaşi.

Talmaciu, M., Percic, S., & Manolescu, I. T. (2023). The Boomerang Effect of Corporate Governance on Public Management – Realities from Romanian Academic Environment. In: C. T. Roman, M. Georgescu, M. Asandului, A. C. Sîrbu (Eds.). *Business Education for a Better World. Conference Proceedings of the XIIIth International Conference Globalization and Higher Education in Economics and Business Administration GEBA* (pp. 349-369), Iași: Universităţii „Alexandru Ioan Cuza" din Iași.

Van Noordt, C., & Misuraca, G. (2022). Artificial intelligence for the public sector: results of landscaping the use of AI in government across the European Union. *Government information quarterly*, *39*(3), 101714. https://doi.org/10.1016/j.giq.2022.101714

Wilson, C., & Van Der Velden, M. (2022). Sustainable AI: An integrated model to guide public sector decision-making. *Technology in Society*, *68*, 101926. https://doi.org/10.1016/j.techsoc.2022.101926

Yawar, M. E., & Hakimi, M. Q. (2025). The Impact of Robots and Artificial Intelligence on Human Resources in the Future. *Global Spectrum of Research and Humanities*, *2*(1), 87-97. https://doi.org/10.69760/gsrh.010120250014

Balkans JETSS (2025) 2: 90-102
DOI: https://doi.org/10.31410/Balkans.JETSS.2025.8.2.90-102

Review Article

Unpacking Cybersecurity in Financial Research: A Systematic Mapping Approach

Leoš Šafár[1]
Michal Mešťan[2]
Jakub Sopko[3]

Received: October 10, 2025 / Revised: December 29, 2025 /
Accepted: December 30, 2025 / Published: December 30, 2025

Abstract: *This study presents a comprehensive bibliometric and content analysis of cybersecurity research in financial institutions covering the period from 2000 to 2025. By applying the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) methodology, we systematically identified and screened 2,005 peer-reviewed articles indexed in the Web of Science Core Collection. The review focuses on cybersecurity within the financial system, particularly in the banking and insurance sectors, and maps research trends, innovation gaps, and thematic evolutions in this domain. Thematic analysis on the basis of keyword co-occurrence identified three distinct clusters. The first cluster focuses on cybersecurity and artificial intelligence, including keywords such as cyber security, phishing, malware, and machine learning, highlighting AI-driven threat detection and digital defense mechanisms. The second cluster focuses on institutional risk management and information security, emphasizing governance, authentication, and systemic controls within financial institutions. The third cluster involves emerging digital technologies, such as blockchain, cloud computing, and FinTech, underscoring the technological innovations that shape financial services and data protection frameworks. Despite growing research output and increasing attention to cyber risk from both academic and regulatory perspectives, the field remains fragmented, with limited empirical evaluation of cybersecurity investment effectiveness. By combining PRISMA guidelines with bibliometric mapping, this study offers a structured and interdisciplinary overview of cybersecurity research in financial institutions. It highlights key research frontiers and calls for deeper empirical inquiry, particularly on systemic financial risk, cross-sector governance, and cybersecurity policy design.*

Keywords: *Cybersecurity, Financial institutions, PRISMA, Bibliometric analysis, Operational risk, Cyber resilience*

JEL Classification: G21 · M15 · O33

✉ leos.safar@tuke.sk
[1] Technical University of Košice, Faculty of Economics, Department of Banking and Investment, Němcovej 32, 040 01 Košice, Slovakia
[2] Matej Bel University in Banská Bystrica, Faculty of Economics, Department of Finance and Accounting, Tajovského 10, 975 90 Banská Bystrica, Slovakia
[3] Technical University of Košice, Faculty of Economics, Department of Banking and Investment, Němcovej 32, 040 01 Košice, Slovakia

1. INTRODUCTION

The accelerating digital transformation of financial systems has redefined how banks, insurers, and related institutions operate, interact with customers, and manage risk. Developments such as mobile banking, algorithmic trading, blockchain-based settlement, and cloud computing have increased operational efficiency and broadened access to services, yet they have also expanded the attack surface for malicious actors. The frequency, scale, and sophistication of cyber threats, including phishing, ransomware, advanced persistent threats (APTs), and large-scale data breaches that jeopardize both institutional viability and systemic stability, have increased (Kemp et al., 2021). The International Monetary Fund (IMF) estimates that severe cyber incidents could cost the global banking sector nearly 9% of its annual profits, underscoring the economic magnitude of the challenge (Bouveret, 2018). The multidimensional nature of cyber risk, which spans technical vulnerabilities, governance failures, and regulatory shortcomings, complicates mitigation efforts. Empirical evidence shows that governance structures, such as the presence of a Chief Risk Officer or dedicated IT committees, can moderate the effectiveness of cybersecurity disclosure and influence bank performance (Elsayed et al., 2024). In parallel, emerging technologies such as artificial intelligence (AI), blockchain, and FinTech applications are transforming both security capabilities and threat vectors (Brando et al., 2022; Sethi et al., 2025; Woods & Böhme, 2021). While FinTech innovations can increase efficiency and expand financial inclusion, they also introduce operational complexities and, in some contexts, short-term profitability pressures due to high implementation costs and intensified competition (Elmahdy et al., 2025; Sethi et al., 2025).

Cybersecurity threats in the banking sector have risen sharply alongside the rapid digitalization, with the IMF reporting a post-COVID-19 doubling of attacks and incidents in finance, which now comprise nearly one-fifth of all cases. Severe cyber incidents now cost up to $2.5 billion annually, with additional reputational and operational losses (International Monetary Fund, 2024). Sulong et al. (2025) analysed U.S. bank data during the period 1998-2018 and showed that heightened cybersecurity risk is positively associated with increased bank risk-taking, especially in smaller, financially vulnerable institutions and those under greater competitive pressure or facing high deposit withdrawals. Factors such as disclosure tone, IT investment intensity, and goodwill condition this relationship, underscoring the complex interplay between cyber risk exposure and strategic behavior in the banking sector.

The policy environment is also evolving. Regulatory bodies, including the Basel Committee, the IMF, and the OECD, have integrated cybersecurity into prudential frameworks, advocating risk-based oversight and systemic resilience (Aldasoro et al., 2020; OECD, 2022; Ravikumar, 2025). PRISMA-based reviews in related domains, such as sustainable finance and central bank digital currencies (CBDCs), highlight the methodological value of structured evidence synthesis in identifying thematic trends, regulatory priorities, and innovation gaps (Galletta et al., 2024; Prodan et al., 2024). The PRISMA-based and bibliometric reviews examine fintech innovation, digital finance, operational risk, or central bank digital currencies, but cybersecurity is typically treated as a supporting topic within these broader themes. In contrast, this study offers a systematic mapping focused exclusively on cybersecurity in financial institutions, with emphasis on banking.

Addressing this gap requires an integrative approach that combines bibliometric analysis with thematic synthesis to uncover research frontiers and fragmentation patterns. This study applies the PRISMA methodology to 2,005 Web of Science–indexed publications (2000–2025) to map the intellectual and conceptual landscape of cybersecurity in financial institutions, with an emphasis on banking and insurance. By identifying clusters of research activity, tracing the evolution of

themes, and pinpointing underexplored areas, we contribute a sector-specific evidence base that can inform academic inquiry, regulatory design, and institutional strategies aimed at strengthening cyber resilience.

In this study, the terms “cybersecurity” and “cyber security” are used interchangeably.

This study addresses four research questions:

- **Research Question One (RQ1):** Which journals are the most influential in the field of cybersecurity research within financial institutions?
- **Research Question Two (RQ2):** What are the major themes and topics emerging from cybersecurity research in the financial sector?
- **Research Question Three (RQ3):** Which of the most influential articles shape cybersecurity research in financial institutions?
- **Research Question Four (RQ4):** What are the future research directions for cybersecurity in the financial sector?

The remainder of this article is structured as follows. Section 2 reviews the relevant literature, positioning this study within existing research on cybersecurity in financial institutions. Section 3 outlines the design and methodological approach, including the PRISMA flowchart and bibliometric techniques employed. Section 4 presents the results of the science mapping and thematic analysis. Section 5 discusses the implications of the findings and proposes directions for future research. Section 6 concludes the paper by summarizing key contributions and highlighting policy and practice implications.

2. LITERATURE REVIEW

Research on cybersecurity in financial institutions has broadened from technical safeguards toward governance, disclosure, and systemic-risk perspectives as digitalization has accelerated. Foundational texts document rising costs and the need for coherent frameworks, whereas policy reports emphasize resilience and coordinated oversight across the sector. These strands frame three concise themes our mapping will interrogate (Bouyon & Krause, 2018; Daimi & Peoples, 2021).

2.1. Cybersecurity Threat Landscape in Finance

Recent reports underscore that cyber incidents can jeopardize financial stability, not merely firm-level operations. The Bangladesh Bank/SWIFT case is widely cited as a systemic wake-up call, and international bodies have since warned that a major cyberattack could precipitate a financial crisis (Maurer & Nelson, 2021). Foundational overviews also stress that, despite proliferating standards and spending, the cost of attacks continues to rise, requiring continual adaptation of defenses (Daimi & Peoples, 2021). At the macro level, the IMF’s Global Financial Stability Report (International Monetary Fund, 2024) treats cyber risk as an emerging concern for macrofinancial stability, placing it firmly on the financial stability agenda.

2.2. Governance, Regulation, and Risk Management

Policy analysis recommends shifting emphasis from ad hoc protection to resilience and clarifies three pillars - governance, risk management, and capability, alongside concrete needs such as convergent incident-reporting taxonomies and better statistics (Bouyon & Krause, 2018). Sectoral overviews highlight that threats and under-reporting expanded markedly in financial services,

motivating organization-wide risk management rather than purely technical fixes (Taplin, 2016). At the firm level, empirical evidence links cybersecurity disclosure to real economic relationships with major customers, indicating that external stakeholders value a stronger cybersecurity posture (Nelson & Wang, 2024). Complementary practice guidance urges finance and accounting professionals to move beyond passive awareness toward structured, risk-based programs aligned with the NIST Cybersecurity Framework (Pendley, 2018).

2.3. Thematic Trends and Research Gaps

Across the literature, three themes recur: (i) evolving attacks surface with digital transformation; (ii) enterprise risk management and disclosure; and (iii) policy and supervisory architectures for resilience. Daimi and Peoples (2021) map enterprise risk frameworks and identity/operations management as organizing concepts for research and practice. The broader AI-driven digital economy heightens exposure and calls for adaptation, reinforcing the need to connect technology, governance, and policy in financial-sector settings (Makarenko et al., 2023; Pacelli, 2025). Moreover, policy analysts describe persistent fragmentation and coordination gaps, and our science-mapping approach is designed to surface and structure these gaps (Maurer & Nelson, 2021).

3. DESIGN AND METHOD

This study employs the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) methodology to ensure a transparent, rigorous, and replicable review process. Although PRISMA originated in health sciences (Liberati et al., 2009; Moher et al., 2009), it has been widely adopted in management, finance, and information systems research for structuring systematic reviews (Cooper et al., 2018; Tranfield et al., 2003). The method is particularly suited to synthesizing and mapping large bodies of academic literature, enabling the identification of thematic trends, intellectual structures, and research gaps. We use updated PRISMA guidelines (Page et al., 2021) and combine them with science mapping techniques implemented in the bibliometrix package for R (Aria & Cuccurullo, 2017). This integration allows for both qualitative synthesis and quantitative bibliometric analysis, providing a comprehensive view of how cybersecurity research in financial institutions has evolved over time. Following the structured approach of PRISMA, our review proceeded through four sequential phases (more detailed in Figure 2).

1. *Identification:* The initial search was conducted in the Web of Science Core Collection database, which was selected for its comprehensive indexing of peer-reviewed literature and compatibility with bibliometric tools. Search terms were developed to capture a broad scope of studies related to cybersecurity in financial institutions, including combinations of "cybersecurity" OR "information security" OR "cyber risk" AND "bank*" OR "insurance" OR "financial institution*". The search was limited to journal articles, conference papers, and review papers published in English between January 2000 and June 2025 (see more details in Table 1).
2. *Screening:* All records were reviewed by title, abstract, and keywords to assess relevance. Studies unrelated to cybersecurity in financial contexts, such as those focusing on non-financial industries or general IT security without sectoral application, were removed. Duplicate entries were also excluded during this stage.
3. *Eligibility:* The remaining records were examined to confirm compliance with the inclusion criteria: (i) direct focus on cybersecurity, information security, or cyber risk; (ii) explicit relevance to financial institutions (banking, insurance, or related financial services); and (iii) publication in peer-reviewed outlets. Non-English records, incomplete studies, or those lacking methodological transparency were excluded.

4. *Inclusion:* After applying the eligibility criteria, the final dataset comprised 2,005 publications. Bibliographic data were exported from Web of Science (WoS) in plain text format and processed with the bibliometrix package. The analyses included performance metrics, co-authorship networks, co-occurrence mapping of author keywords, and thematic evolution analysis. These techniques facilitated the identification of influential authors, journals, articles, and research clusters, directly addressing the study's research questions.

By combining the structured PRISMA protocol with bibliometric science mapping, this study ensures methodological transparency, replicability, and analytical depth, offering a robust synthesis of two and a half decades of cybersecurity research in financial institutions.

4. RESULTS

Figure 1 shows the annual distribution of publications on cybersecurity in financial institutions from 2000 to June 2025. Research output was minimal in the early 2000s, with a noticeable spike in 2008, followed by a decline and gradual growth until 2013. A sustained upward trend began in 2014, accelerating sharply between 2016 and 2019 and reaching its peak in 2024, with over 220 publications. The lower count in 2025 reflects mid-year data collection rather than an actual drop in research activity.

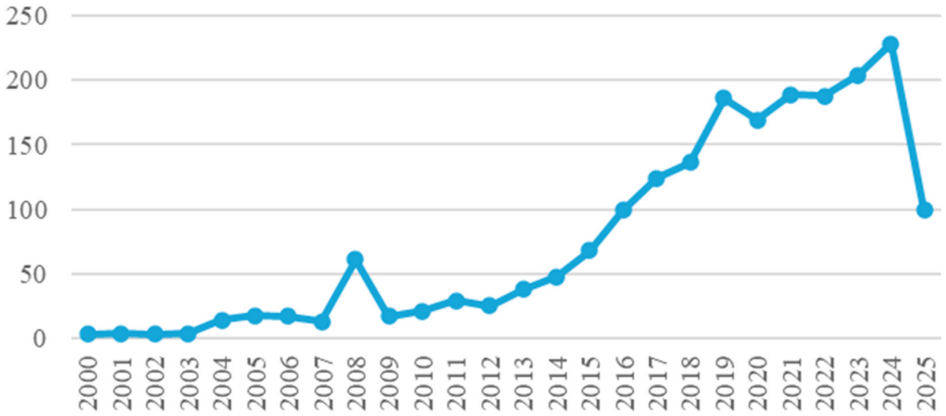

Figure 1. Total number of publications between 2000 and June 2025
Source: Own processing

Table 1 summarizes the PRISMA model applied in this study, outlining the review design, protocol, eligibility criteria, and search strategy. We combined two thematic dimensions, cybersecurity and the financial sector (Uddin et al., 2020), using predefined Boolean codes in the Web of Science Core Collection to ensure comprehensive and replicable coverage of relevant literature.

Figure 2 presents the PRISMA flow diagram summarizing the selection process for the systematic review. The initial keyword search in the Web of Science Core Collection returned 86,936 records for cybersecurity-related terms and 587,860 for financial-sector terms. When these two thematic dimensions were combined via Boolean logic, the search yielded 2,141 documents, which formed the initial dataset for screening and analysis. Following the PRISMA 2020 guidelines (Page et al., 2021) to ensure transparency and replicability, seven items, such as retracted publications, book reviews, and meeting abstracts, were removed, leaving 2,134 records. Further screening excluded three items indexed only in the Arts & Humanities Citation Index and 32 articles not

written in English, resulting in 2,099 records. The research area filter removes 94 unrelated items (e.g., works in religion or thermodynamics), producing the final dataset of 2,005 publications used for bibliometric and thematic analyses.

Table 1. PRISMA model for systematic literature review

Study design	The study conducts a literature review that synthesizes existing research through a rigorous, clearly defined, and transparent step-by-step process.
Review protocol	To reduce the risk of "biased post hoc decisions in review methods" (Liberati et al., 2009), the search criteria and corresponding keywords were defined in advance.
Eligibility criteria	The review included only articles published in peer-reviewed journals, conference proceedings, reviews, and book chapters. Studies were retrieved from the Web of Science electronic database using search codes established by the authors, and the resulting bibliometric data were subsequently mapped and clustered.
Publication type included	Peer-reviewed journals, conference proceedings, reviews, and book chapters from the Web of Science database
Publication timeframe	2000 – June 2025
Language	English
Search strategy	We have used two thematic areas: cybersecurity and financial sector. We selected the following codes to search in the WoS database: ALL= "cybersecurity" OR "cyber security" OR "cyber-security" OR "cyber risk" OR"cyber attack", "cyber-attack" OR "cyber threat" OR "cyber harm", "information security" OR "data breach" OR "IT security" OR "cyber resilience" OR "digital security" AND "bank*" OR "financial institution*" OR "financial sector" OR "insurance compan*" OR "commercial bank*" OR "central bank*" OR "fintech" OR "financial service*" OR "banking vulnerability"

Source: Own processing

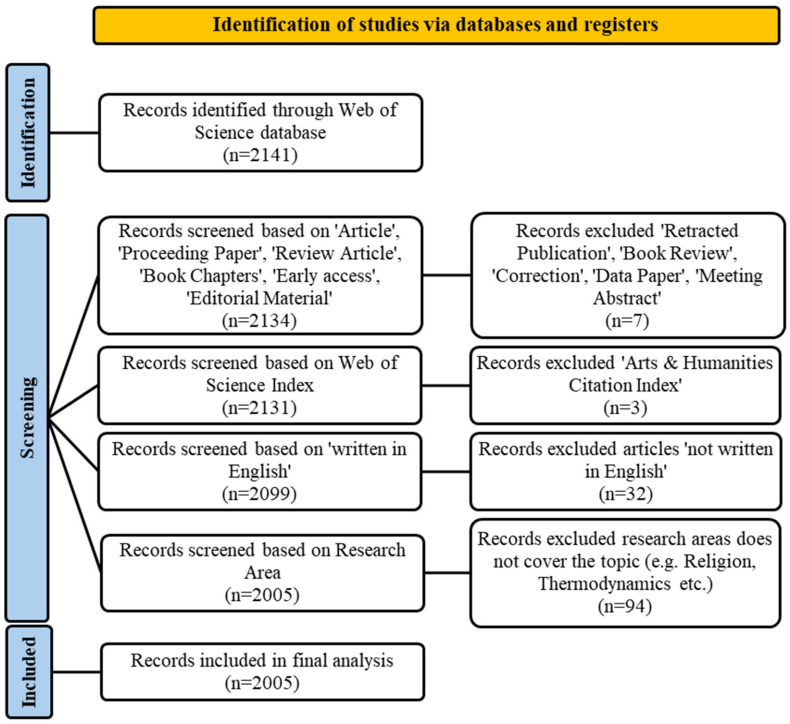

Figure 2. PRISMA flowchart for the systematic literature review

Source: Own processing

Table 2 lists the most productive and influential journals publishing on cybersecurity in financial institutions (RQ1). Computers & Security leads in output with 45 publications and an h-index of 21, followed by IEEE Access (44 publications; h-index 15) and Information and Computer Security (21 publications; h-index 10). Citation metrics such as the g-index and m-index confirm the consistent impact of these outlets over time, positioning them as core publication venues for the field. The *h*-index measures a journal's impact by counting how many articles, *h*, have at least *h* citations. The *g*-index is a similar metric that gives more weight to highly cited papers. It identifies the top *g* articles that have received a total of at least g^2 citations. To compare journals of different ages, the *m*-index is a useful adjustment. It divides the *h*-index by the number of years the journal has been published.

Table 2. Distribution of publications across influential journals

Sources	No. of articles	Total citations	*h*_index	*g*_index	*m*_index	Publication year start
Computers & Security	45	1663	21	40	0.84	2001
IEEE Access	44	1308	15	36	1.67	2017
Information and Computer Security	21	365	10	19	0.91	2015
Journal of Information Security and Applications	19	736	10	19	1.11	2017
Information Security Journal	18	73	6	7	0.35	2009
International Journal of Information Security and Privacy	16	84	4	9	0.21	2007
International Journal of Information Security	15	110	7	10	0.41	2009
Sensors	14	168	9	12	0.53	2009
Scientific Reports	13	183	5	13	0.56	2017
Applied Sciences - Basel	11	104	5	10	0.63	2018
Financial and Credit Activity - Problems of Theory and Practice	11	16	2	2	0.33	2020

Source: Own processing

The author-keyword co-occurrence analysis (Figure 3) reveals three distinct thematic clusters shaping the field (RQ2):

1. Technological Threat Detection and AI Integration (Red cluster)
    a. This includes "cybersecurity," "artificial intelligence," "machine learning," "phishing," "fraud," and "malware," indicating strong research activity on AI-driven threat identification, anomaly detection, and predictive modeling. The high density of links in this cluster demonstrates tight conceptual integration, reflecting the sector's increasing reliance on data-driven security tools. From a policy perspective, the prominence of this cluster emphasizes the importance of regulatory frameworks that address algorithmic risk alongside operational resilience.
2. Governance, Risk Management, and Institutional Controls (Blue cluster)
    a. Centered on "information security," "risk management," and "authentication," this cluster reflects research that integrates cybersecurity with enterprise risk frameworks, authentication protocols, and systemic governance measures within financial institutions. Its centrality also underscores that cybersecurity is no longer considered as an IT function but rather one of the main pillars of financial stability and regulatory compliance, which are reinforced by regulatory initiatives that embed cybersecurity risk into capital adequacy, governance standards, and supervisory stress testing.
3. Emerging Digital Technologies and Privacy Concerns (Green cluster)
    a. Keywords such as "blockchain," "FinTech," "cloud computing," "Internet of Things," and "privacy" appear here, signaling an active stream of research exploring both the

opportunities and vulnerabilities posed by technological innovations in finance. It suggests the need for a balance between innovation incentives and data protection, interoperability, and coordination between the different sectors in cybersecurity.

The keyword frequency data reinforce this structure, with "cyber security" (550 occurrences), "information security" (202), "blockchain" (97), "machine learning" (93), "privacy" (59), "fintech" (59), and "phishing" (46) ranking among the most frequent terms in the dataset.

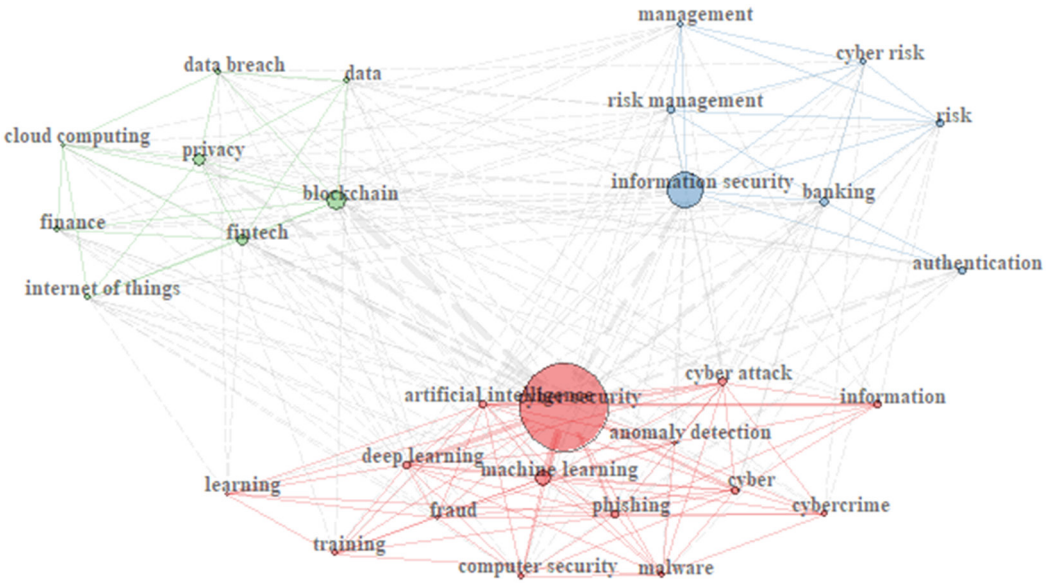

Figure 3. Keyword co-occurrence network diagram
Source: Own processing

The thematic evolution map (Figure 4) shows how these topics have shifted over time, segmented into three periods (RQ4): 2000–2015, 2016–2020, and 2021–2025. The first cut-off point (year 2015) precedes the 2016 Bangladesh Bank SWIFT cyber heist (Maurer & Nelson, 2021), which marked a turning point in industry awareness of systemic financial cyber threats. Early themes (during the period 2000–2015), such as "internet banking," "integrity," and "cyber security," gave way in 2016–2020 to more specialized topics, including "biometric authentication," "information security culture," "machine learning," and "malware." The second break (year 2020) aligns with the COVID-19 pandemic, which accelerated digital transformation and remote financial services, pushing "cyber risk," "cybersecurity," and "security" to prominence in the 2021–2025 period. This evolution highlights a sector-wide shift from foundational security measures to advanced, AI-supported risk mitigation and resilience strategies. The thematic mapping results (Figure 5) classify "privacy" as a motor theme, suggesting that it is both central and well-developed; "security" as a basic theme, indicating foundational relevance; "risk" as a niche theme, associated with specialized but significant inquiry; and "algorithm" in the emerging/declining quadrant, highlighting a need for renewed exploration of AI governance and resilience.

Table 3 lists the most locally cited articles within the dataset, highlighting those that have exerted a disproportionate influence on the research field (RQ3). Biener et al. (2015), examining the insurability of cyber risk, led with 29 local citations, framing much of the debate around financial risk transfer and resilience.

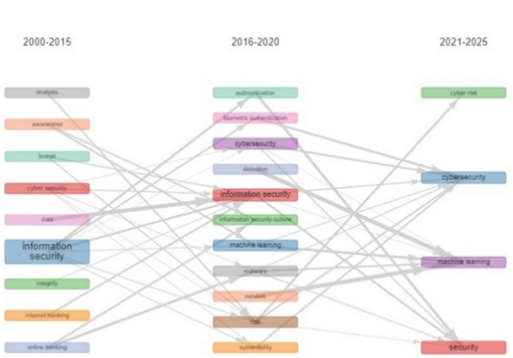

Figure 4. Thematic evolution map
Source: Own processing

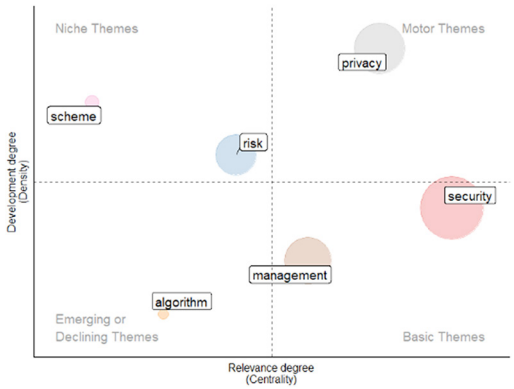

Figure 5. Thematic Keywords Plus map
Source: Own processing

Table 3. Co-citation analysis

Authors & Year	Topic	Cit.
Biener et al. (2015)	Insurability of cyber risk: An empirical analysis.	29
Davis (1989)	Perceived usefulness, perceived ease of use, and user acceptance of information technology	29
Gordon & Loeb (2002)	The economics of information security investment	23
Herath & Rao (2009)	Protection motivation and deterrence: a framework for security policy compliance in organizations	22
Cavusoglu et al. (2004)	The effect of internet security breach announcements on market value: Capital market reactions for breached firms and internet security developers	19
Kamiya et al. (2021)	Risk management, firm reputation, and the impact of successful cyberattacks on target firms.	19

Source: Own processing

Gordon and Loeb (2002), with 23 citations, provide a foundational economic model for evaluating investments in information security widely applied in banking and insurance contexts. Together with other studies, these references anchor economic, behavioral, and market-impact perspectives that the field repeatedly builds on.

5. FUTURE RESEARCH DIRECTIONS

The evolution of cybersecurity research in financial institutions, as mapped in this study, reveals several trajectories that warrant deeper exploration. Emerging technological, regulatory, and organizational developments are reshaping the security landscape, creating opportunities for multidisciplinary, empirically grounded inquiry. First, the continued integration of artificial intelligence (AI), machine learning, and deep learning into cyber defense systems opens new avenues for research. While recent work has demonstrated the potential of AI-driven anomaly detection and predictive modelling (Sethi et al., 2025; Woods & Böhme, 2021), further investigation into algorithmic transparency, model governance, and resilience against adversarial attacks in high-stakes financial contexts is needed. The interaction between AI-based security tools and evolving regulatory requirements remains a largely underexplored area. Second, the rapid adoption of blockchain, FinTech, and cloud computing in banking and insurance presents novel

vulnerabilities alongside efficiency gains (Brando et al., 2022; Elmahdy et al., 2025). Future research should examine the design of secure architecture for these technologies, particularly under cross-border operational frameworks, and assess their role in mitigating or exacerbating systemic risk. Third, the increasing salience of privacy as both a regulatory mandate and a competitive differentiator calls for studies that connect technical privacy-enhancing measures with compliance obligations under frameworks such as the GDPR and emerging data localization laws (OECD, 2022). Longitudinal research could evaluate how privacy practices influence customer trust, firm performance, and market stability. Fourth, cyber risk quantification and modelling remain critical challenges. Building on work linking cyber risk to financial risk-taking (Sulong et al., 2025) and economic loss estimates (Bouveret, 2018; International Monetary Fund, 2024), future studies should advance predictive models that integrate macro-financial indicators, interbank connectivity, and behavioural factors. Such models could inform regulatory stress testing and capital adequacy assessments for cyber resilience. Finally, governance and cross-sector coordination merit deeper empirical investigation. While governance mechanisms such as dedicated IT committees and Chief Risk Officers have been associated with better cybersecurity outcomes (Elsayed et al., 2024), the effectiveness of these structures in different regulatory and cultural environments remains unclear. Comparative international studies could illuminate best practices for aligning organizational governance with evolving cyber threat landscapes. By addressing these themes, future research can bridge the technical, economic, and policy dimensions of financial-sector cybersecurity, ensuring that research efforts keep pace with the sector's rapidly shifting risk environment.

6. CONCLUSION

This study applied PRISMA-based systematic mapping and bibliometric analysis to 2,005 publications on cybersecurity in financial institutions from 2000 to 2025. By combining structured literature screening with science mapping via the bibliometrix package, we identified influential journals, highly cited works, major thematic clusters, and evolving research trends. The findings reveal a diverse but interconnected research landscape with three primary thematic domains: (i) AI-driven threat detection and technological safeguards, (ii) governance and institutional risk management, and (iii) emerging digital technologies and privacy.

The analysis highlights the increasing academic and policy focus on cybersecurity as a systemic risk factor for the financial sector, particularly following major inflection points such as the 2016 Bangladesh Bank cyber heist and the 2020 COVID-19 pandemic. These events have shifted the research frontier toward integrating advanced technologies with regulatory frameworks, enhancing governance structures, and addressing privacy and data protection in a globalized, digitally dependent financial system.

From a practical perspective, this study contributes sector-specific evidence-based that can inform institutional strategy and policy design aimed at strengthening cyber resilience. The results are relevant to regulators seeking to design adaptive oversight mechanisms, financial institutions aiming to integrate cybersecurity into enterprise risk management, and technology providers developing sector-specific security solutions.

This study has several limitations that provide avenues for additional research. First, the analysis was restricted to publications indexed in the Web of Science Core Collection, potentially omitting relevant work from other databases. Second, the focus on English-language publications may introduce a language bias, underrepresenting perspectives from non-English-speaking regions.

Third, bibliometric methods privilege citation-based influence, which may not always correlate with practical impact or emerging niche research. Finally, the dataset reflects research activity up to mid-2025, meaning that very recent developments may not yet be fully represented in the available literature.

Despite these constraints, this study provides a transparent and replicable framework for understanding the intellectual and thematic structure of cybersecurity research in financial institutions. Mapping the evolution of the field and identifying thematic gaps offers both a consolidated reference for current research and a foundation for future inquiry.

Acknowledgment

This work was partially supported by the MINEDU of the Slovak Republic [grant No. 1/0638/25] and the Slovak Research and Development Agency under Contract no. VV-MVP-24-0272.

References

Aldasoro, I., Gambacorta, L., Giudici, P., & Leach, T. (2020). Operational and cyber risks in the financial sector. *BIS Working Papers*, No 840.

Aria, M., & Cuccurullo, C. (2017). bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics*, *11*(4), 959-975. https://doi.org/10.1016/j.joi.2017.08.007

Biener, C., Eling, M., & Wirfs, J. H. (2015). Insurability of cyber risk: An empirical analysis. *The Geneva Papers on Risk and Insurance - Issues and Practice*, *40*(1), 131-158. https://doi.org/10.1057/gpp.2014.19

Bouveret, A. (2018). Cyber Risk for the Financial Sector: A Framework for Quantitative Assessment. *IMF Working Papers, 18*(143), 1. https://doi.org/10.5089/9781484360750.001

Bouyon, S., & Krause, S. (2018). *Cybersecurity in finance: Getting the policy mix right*. Rowman & Littlefield.

Brando, D., Kotidis, A., Kovner, A., Lee, M., & Schreft, S. L. (2022). Implications of Cyber Risk for Financial Stability. *FEDS Notes, 2022.0*(3077.0). https://doi.org/10.17016/2380-7172.3077

Cavusoglu, H., Mishra, B., & Raghunathan, S. (2004). The effect of internet security breach announcements on market value: Capital market reactions for breached firms and internet security developers. *International Journal of Electronic Commerce*, *9*(1), 70-104. https://doi.org/10.1080/10864415.2004.11044320

Cooper, C., Booth, A., Varley-Campbell, J., Britten, N., & Garside, R. (2018). Defining the process to literature searching in systematic reviews: a literature review of guidance and supporting studies. *BMC Medical Research Methodology*, *18*(1), 85. https://doi.org/10.1186/s12874-018-0545-3

Daimi, K., & Peoples, C. (Eds.). (2021). *Advances in Cybersecurity Management*. Springer. https://doi.org/10.1007/978-3-030-71381-2

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly, 13*(3), 319-340. https://doi.org/10.2307/249008

Elmahdy, A. H. A. M., Abdelkader, M. T. K. M., & Shaker, M. A. M. (2025). Bridging the nexus between Fintech, operational efficiency and banks profitability: the moderating role of bank size. *Future Business Journal*, *11*(1). https://doi.org/10.1186/s43093-025-00478-x

Elsayed, D. H., Ismail, T. H., & Ahmed, E. A. (2024). The impact of cybersecurity disclosure on banks' performance: the moderating role of corporate governance in the MENA region. *Future Business Journal*, *10*(1). https://doi.org/10.1186/s43093-024-00402-9

Galletta, S., Mazzù, S., Naciti, V., & Paltrinieri, A. (2024). A PRISMA systematic review of greenwashing in the banking industry: A call for action. *Research in International Business and Finance*, *69*, 102262. https://doi.org/10.1016/j.ribaf.2024.102262

Gordon, L. A., & Loeb, M. P. (2002). The economics of information security investment. *ACM Transactions on Information and System Security*, *5*(4), 438-457. https://doi.org/10.1145/581271.581274

Herath, T., & Rao, H. R. (2009). Protection motivation and deterrence: a framework for security policy compliance in organisations. *European Journal of Information Systems*, *18*(2), 106-125. https://doi.org/10.1057/ejis.2009.6

International Monetary Fund. (2024). *Global Financial Stability Report: The Last Mile: Financial Vulnerabilities and Risks*. Washington, DC, April. https://doi.org/10.5089/9798400257704.082

Kamiya, S., Kang, J.-K., Kim, J., Milidonis, A., & Stulz, R. M. (2021). Risk management, firm reputation, and the impact of successful cyberattacks on target firms. *Journal of Financial Economics*, *139*(3), 719-749. https://doi.org/10.1016/j.jfineco.2019.05.019

Kemp, S., Buil-Gil, D., Moneva, A., Miró-Llinares, F., & Díaz-Castaño, N. (2021). Empty streets, busy internet: A time-series analysis of cybercrime and fraud trends during COVID-19. *Journal of Contemporary Criminal Justice*, *37*(4), 480-501. https://doi.org/10.1177/1043986211027986

Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gotzsche, P. C., Ioannidis, J. P. A., Clarke, M., Devereaux, P. J., Kleijnen, J., & Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *BMJ, 339*(jul21 1), b2700-b2700. https://doi.org/10.1136/bmj.b2700

Makarenko, E. N., Vovchenko, N. G., & Tishchenko, E. N. (2023). *Technological Trends in the AI Economy. Smart Innovation, Systems and Technologies.* Springer Nature. https://doi.org/10.1007/978-981-19-7411-3

Maurer, T., & Nelson, A. (2021). The global cyber threat. *Finance & Development*, *58*(1), 24-27.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & the PRISMA Group*. (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *Annals of Internal Medicine, 151*(4), 264-269. https://doi.org/10.7326/0003-4819-151-4-200908180-00135

Nelson, A., & Wang, S. (2024). The importance of cybersecurity disclosures in customer relationships. *Journal of Corporate Accounting & Finance*, *35*(3), 66-74. https://doi.org/10.1002/jcaf.22695

OECD. (2022). *OECD Policy Framework on Digital Security*, OECD Publishing, Paris, https://doi.org/10.1787/a69df866-en

Pacelli, V. (2025). *Systemic Risk and Complex Networks in Modern Financial Systems* (p. 412). Springer Nature. https://doi.org/10.1007/978-3-031-64916-5

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*, n71. https://doi.org/10.1136/bmj.n71

Pendley, J. A. (2018). Finance and accounting professionals and cybersecurity awareness. *Journal of Corporate Accounting & Finance*, *29*(1), 53-58. https://doi.org/10.1002/jcaf.22291

Prodan, S., Konhäusner, P., Dabija, D.-C., Lazaroiu, G., & Marincean, L. (2024). The rise in popularity of central bank digital currencies. A systematic review. *Heliyon*, *10*(9), e30561. https://doi.org/10.1016/j.heliyon.2024.e30561

Ravikumar, R. (2025). Strengthening Cybersecurity: Lessons from the Cybersecurity Survey. *Technical Notes and Manuals, 2025*(006), 1. https://doi.org/10.5089/9798400296864.005

Sethi, M., Bohra, N. S., Johri, A., & Asif, M. (2025). Emerging dimensions in Fintech: Insights from bibliometric analysis. *Digital Business*, *5*(1), 100113. https://doi.org/10.1016/j.digbus.2025.100113

Sulong, Z., Fuszder, M. H. R., Abdullah, M., & Abakah, E. J. A. (2025). Cybersecurity risk and bank risk-taking. *Journal of Behavioral and Experimental Finance*, 101080. https://doi.org/10.1016/j.jbef.2025.101080

Taplin, R. (Ed.). (2016). *Managing Cyber Risk in the Financial Sector*. Routledge, Taylor & Francis Group. https://doi.org/10.4324/9781315675930

Tranfield, D., Denyer, D., & Smart, P. (2003). Towards a methodology for developing evidence-informed management knowledge by means of systematic review. *British Journal of Management*, *14*(3), 207-222. https://doi.org/10.1111/1467-8551.00375

Uddin, M. H., Ali, M. H., & Hassan, M. K. (2020). Cybersecurity hazards and financial system vulnerability: a synthesis of literature. *Risk Management*, *22*(4), 239-309. https://doi.org/10.1057/s41283-020-00063-2

Woods, D. W., & Böhme, R. (2021). SoK: Quantifying cyber risk. In *2021 IEEE Symposium on Security and Privacy (SP)* (pp. 211-228). IEEE. https://doi.org/10.1109/SP40001.2021.00053

Balkans JETSS (2025) 2: 103-113
DOI: https://doi.org/10.31410/Balkans.JETSS.2025.8.2.103-113

Review Article

Partners or Threats? The Hidden Dynamics of AI Adoption in the Workplace

Žikica Milošević[1]
Andrea Ivanišević[2]
Katarina Stojanović[3]

Received: October 23, 2025 / Revised: December 23, 2025 /
Accepted: December 30, 2025 / Published: December 30, 2025

Abstract: *Artificial Intelligence (AI) is rapidly reshaping organisational work, yet its promised benefits often conceal deeper psychological and managerial tensions. This study aims to examine why employees frequently hide their use of AI, how organisational culture shapes these behaviours, and how generational differences influence attitudes toward AI-driven productivity. While companies promote AI as a tool for efficiency and creativity, many workers fear that admitting AI use will lead either to increased workloads or to job displacement. This creates a paradox in which AI is experienced less as a supportive partner and more as a silent competitor or monitoring device. As efficiency gains are reinvested into higher output expectations, anxiety, mistrust, and concealment become common workplace strategies. The paper also analyses how Generation Z responds to this "productivity trap" by simultaneously embracing AI while seeking careers in less automatable fields that promise stability, dignity, and work-life balance. The findings highlight the need for organisations to reframe AI adoption as a collaborative human-machine partnership, supported by transparent communication, ethical management, and redefined performance metrics that protect employee well-being while enabling innovation.*

Keywords: *AI, Workplace, Generation Z, Threats, Partnership, Management*

JEL Classification: J24 · M15 · M54

✉ zikica_m@yahoo.com
[1] Faculty of Technical Sciences Novi Sad, Trg Dositeja Obradovića 6, Novi Sad, Serbia
[2] Faculty of Technical Sciences Novi Sad, Trg Dositeja Obradovića 6, Novi Sad, Serbia
[3] Faculty of Contemporary Arts, Belgrade; Faculty of Economics and Engineering Management, University Business Academy in Novi Sad, Serbia

1. INTRODUCTION

The rapid diffusion of Artificial Intelligence (AI) across workplaces has marked one of the most significant transformations of the 21st century. Once regarded as a futuristic concept, AI has now become a tangible element of everyday professional life - embedded in email filters, predictive analytics, recruitment software, and even creative tools such as text and image generators. This technological shift is often portrayed in corporate narratives as both inevitable and benevolent: a force that promises efficiency, reduces routine work, and liberates human creativity (Davenport & Mittal, 2023; McAfee & Brynjolfsson, 2017). AI, therefore, presents a brand new general-purpose technology with potentially huge impacts on workers and organisations (Farrell et al., 2025). This will be a transformative factor, making nearly every sector of the economy more efficient (Acemoglu, 2025) and less biased (Page & Kallapur, 2025). However, beneath this optimistic discourse lies a growing tension between the promise of AI and the perception of those who must coexist with it. Recent studies reveal that many employees experience AI not as an empowering tool but as a source of anxiety, control, and mistrust (Haenlein et al., 2019; Rahwan et al., 2019). Instead of serving as a neutral assistant, AI systems are often embedded within managerial frameworks that measure performance, predict behaviour, and even inform promotion or termination decisions. This creates a psychological ambivalence: employees are encouraged to "work smarter", yet often feel surveilled, forced to do more for the same period of time, or even replaceable. As Grasiaswaty (2025) notes, the discourse of automation conceals a deeper managerial logic - one that seeks to intensify work through algorithmic monitoring rather than alleviate it. A striking manifestation of this paradox can be seen in what researchers describe as the "AI concealment effect" (Kellogg et al., 2020). Many workers use AI tools privately but hesitate to disclose their reliance on them. The reasons are predominantly organisational rather than technical: fear of job loss, fear of managerial over-expectation, or fear of being perceived as less capable (and thus qualified to become obsolete). In this sense, AI becomes not a symbol of progress, but of precariousness. The workplace turns into a paradoxical arena where employees hide their technological proficiency to safeguard their professional stability. Generational differences further illuminate this phenomenon. Generation Z, entering the workforce during the post-pandemic digital acceleration, approaches AI with both confidence and caution (Deloitte, 2025; Schroth, 2019). Since they are projected to make up 74% of the global workforce by 2030, they are a formidable force (Deloitte, 2025). Digital nativity allows them to use AI creatively and intuitively, yet they are acutely aware of its double-edged nature. Unlike older cohorts, many Gen Z workers reject the glorification of overwork and the "always-on" mentality inherited from the early '00s (Fuchs et al., 2024). Instead, they value psychological safety, work-life balance, and meaningful autonomy - factors that traditional management models often undervalue. Consequently, they gravitate towards roles that cannot be easily automated, such as design, communication, education, and cultural production, viewing them as bastions of human distinctiveness. Some of them even lean towards blue-collar jobs. Namely, 42% of Gen Z adults are currently working in or pursuing a blue-collar or skilled trade job, according to a report from Resume Builder mde in May 2025 (Crist, 2025). Of those, 37% have earned a bachelor's degree, and 25% of them say these jobs are less likely to be replaced by AI (Resume Builder, 2025). As a consequence, qualified workforce from this generational cohort will rather leave the insecure or/and heavily additionally burdened white-collar jobs than comply. This should be a big red flag for HR. However, there are some encouraging data: according to a joint report by MIT Sloan Management Review and the Boston Consulting Group, around 60% of employees perceive artificial intelligence not as a threat to their jobs but as a "co-worker" or collaborator. Moreover, organisations where employees actively derive value from AI are 5.9 times more likely to achieve significant financial benefits than those where employees do not effectively engage with AI tools (Ransbotham et al., 2022). The central problem that this paper addresses is thus twofold. First, it

explores the contradiction between the corporate rhetoric of AI as a productivity enhancer and the lived reality of employees who often perceive it as a source of surveillance or pressure. Second, it examines how generational shifts - especially the values and strategies of Generation Z - might catalyse a more human-centred redefinition of AI integration in organisational life. In doing so, the research seeks to answer the following questions:

1. How does organisational culture influence employees' willingness to openly use and discuss AI tools?
2. What psychological and managerial mechanisms contribute to the concealment of AI use in the workplace?
3. How does Generation Z challenge or reshape traditional managerial expectations in the context of AI adoption?

By framing AI not merely as a technological innovation but as a socio-economic and cultural phenomenon, this study contributes to the growing discourse on human-machine collaboration, digital trust, and ethical management. It argues that the future of AI in the workplace will depend less on its computational capabilities and more on how organisations choose to reimagine relationships of trust, transparency, and shared value (Chung & Schiff, 2025; Susskind & Susskind, 2015). Unless AI is redefined as a partner rather than a threat, it risks deepening the managerial paradox of modern capitalism - where progress in technology coincides with regression in human well-being. It is worth noticing that this is a conceptual paper, i.e. a literature review with no primary data, since the future research will dig more deeply into the subject.

2. FROM AUTOMATION TO ALGORITHMIC MANAGEMENT – THE RISE OF "TECHNOLOGICAL TAYLORISM"

The emergence of AI-driven technologies has reshaped the fundamental architecture of organisational management. What was once a hierarchy governed by human supervisors is increasingly mediated by algorithmic systems that track, predict, and evaluate performance. This shift has given rise to what Kellogg et al. (2020) call the "new contested terrain of control", in which algorithms act simultaneously as instruments of coordination and surveillance. The early promise of automation - reducing routine labour to allow for creative and strategic engagement - has gradually evolved into a more complex managerial apparatus, often described as algorithmic management (Mateescu & Nguyen, 2019). Algorithmic management refers to the delegation of supervisory tasks - monitoring, scheduling, and evaluating - to software systems rather than human managers (Möhlmann & Zalmanson, 2017). While efficiency gains are substantial, scholars have noted that such systems may also erode autonomy and increase psychological strain (Lee et al., 2015). In many workplaces, AI acts as both a "digital boss" and a "data-driven mirror," reflecting employees' performance back to them with unprecedented precision. Yet this precision often comes at the cost of trust. Grasiaswaty (2025) argues that the rhetoric of efficiency conceals a deeper managerial intention: AI is used not only to streamline work but to intensify it by quantifying every moment of labour. A parallel body of literature suggests that this trend represents a new stage of so-called "technological Taylorism", where algorithmic systems extend the logic of early industrial management into the cognitive and emotional domains of work (Kellogg et al., 2020; Rosenblat, 2018). As a result, AI becomes less a tool of assistance and more an agent of standardisation. Even in creative and knowledge-based industries, where individuality and improvisation are vital, employees increasingly encounter digital metrics of "engagement" or "creativity" imposed by management dashboards.

3. THE PSYCHOLOGICAL LANDSCAPE: TRUST, FEAR, AND CONCEALMENT

While corporate narratives celebrate AI as a "neutral partner in progress", psychological and sociological research paints a more ambivalent picture. Rahwan et al. (2019) describe this as a "machine behaviour paradox"- humans design algorithms but then must adapt to their logic. This adaptation often triggers anxiety, not because of technical difficulty or the error made by the machine (AI outputs), but because of uncertainty about how AI outputs are interpreted by managers. The opacity of algorithmic decision-making further deepens this anxiety (Burrell, 2016; Haenlein et al., 2019). Recent workplace studies reveal that employees frequently conceal their use of AI tools, fearing that transparency could backfire. This "AI concealment effect" demonstrates a misalignment between technological innovation and organisational culture. When employees believe that admitting AI reliance could either signal redundancy or justify increased workloads, the supposed benefits of automation become self-defeating. Instead of enhancing productivity, fear of exposure generates inefficiency through emotional self-censorship and reduced openness. Scholars in organisational psychology have linked this dynamic to the broader theory of psychological safety - a condition where individuals feel free to express ideas or admit mistakes without fear of punishment (Edmondson, 2018). In AI-mediated workplaces, psychological safety depends less on technology itself and more on managerial framing. If AI is positioned as a monitoring device rather than a collaborator, employees internalise a sense of surveillance and control (Huang & Rust, 2018). Conversely, when AI is introduced with transparent communication, ethical guidelines, and participatory training, trust tends to increase (Haenlein et al., 2019). Moreover, the literature suggests that the fear of AI replacement is not only an economic concern but also an existential one. Susskind and Susskind (2015) argue that AI challenges professional identity by questioning what constitutes "uniquely human" labour. For many, the notion of being outperformed by a machine undermines personal dignity and the meaning of work. This explains why concealment may coexist with heavy AI usage: workers rely on AI pragmatically but deny it symbolically to preserve self-worth and their status in the eyes of managers.

4. GENERATIONAL AND CULTURAL DIMENSIONS: GENERATIONS Y AND Z AND THE SEARCH FOR MEANING

The literature on generational behaviour provides important insights into how digital-native cohorts navigate the new AI workplace. Generation Z, born between the mid-1990s and early 2010s, represents the first generation to enter the workforce with full digital fluency and algorithmic literacy (Schroth, 2019). Unlike Millennials (Generation Y), whose early careers were marked by technological adaptation, Gen Z workers have grown up within ecosystems of predictive algorithms, social media analytics, and recommendation systems. This familiarity, however, does not necessarily translate into unconditional acceptance. Deloitte's 2025 Gen Z and Millennial Survey found that among those who already use GenAI at work, Gen Zs and millennials are broadly optimistic about its impact. They believe it has improved the quality of their work (78% of Gen Zs and 82% of millennials), and that it has helped to free up their time and improved their work/life balance (77% of Gen Zs and 79% of millennials). When asked about the top impacts of GenAI on their careers so far, roughly a quarter (28%) of both generations say that GenAI has simplified routine tasks and increased efficiency, improved overall productivity (25% of Gen Zs and 26% of millennials), and enhanced creativity and innovation (25% of Gen Zs and 24% of millennials). Twenty percent of Gen Zs and millennials say that GenAI has freed them up to focus on more strategic work. But they do have some significant concerns. They worry GenAI will eliminate jobs (63% of Gen Zs and 65% of millennials) and make it harder for younger generations to enter the workforce as it automates tasks typically performed by entry-level

workers (61% of Gen Zs and millennials). The prevalence of GenAI is also causing many (66% of Gen Zs and 68% of millennials) to say they will look for job opportunities that they perceive as being safe from GenAI-driven disruption, such as manual labour or skilled trades. This is up from last year, when less than six in 10 Gen Zs (59%) and just over half of millennials (52%) said the same (Deloitte, 2025). This ambivalence reflects what is described as the "productivity trap" - the perception that technology designed to save time ultimately increases pressure to do more. Gen Z's response to this dilemma is both strategic and philosophical. On the one hand, they adopt AI to reduce cognitive load and streamline routine tasks – they want AI to be their "copilot" or an autopilot, not a threat, a competitor, or a servant. On the other hand, they actively seek roles that preserve the human factor: creativity, empathy, and social connection, relegating laborious and dull tasks to their AI partner. This search for meaningful work aligns with broader sociological shifts towards post-material values and human sustainability (Kraus et al., 2019). Gen Z workers, more than any previous generation, frame success not in terms of income or status but in terms of balance, autonomy, and authenticity. Their resistance to the "grind culture" challenges management to redefine performance metrics that value human judgment and ethical reflection. Their tendency for well-being is opposed by the old-school billionaires and company owners like Jeff Bezos, Elon Musk and Reid Hoffman, since Bezos considers any "balance" to be a "tradeoff" (Burleigh, 2025), and the opposing forces will continue to clash in the future. As Chung and Schiff (2025) argue, sustainable AI integration requires a "social contract for automation" that redistributes efficiency gains in ways that protect psychological health and civic dignity. Across these strands of research, one theme is consistent: the social meaning of AI depends not on its algorithms but on the contexts in which they operate. When embedded within exploitative or opaque managerial cultures, AI reinforces asymmetries of power and weakens employee trust. Yet, when aligned with participatory ethics, transparent governance, and human-centred design, AI can genuinely augment rather than replace human capacities. The challenge for organisations, therefore, is not merely technological adoption but cultural translation - reframing AI as a cooperative partner rather than a disciplinary instrument.

5. ORGANISATIONAL CULTURE AND THE OPENNESS TO USE AI TOOLS

The organisational culture in which employees operate has become a crucial determinant of how artificial intelligence (AI) is adopted and discussed within the workplace. AI integration is not solely a technological or managerial issue but a cultural one. As Schein and Schein (2017) argue, culture shapes "the way things are done" in an organisation - what is rewarded, tolerated, or silenced. When applied to AI, this means that trust, transparency, and psychological safety define whether employees will use AI tools openly or conceal their reliance on them. Empirical studies show that innovative and learning-oriented cultures tend to facilitate openness toward AI. For example, Huang and Rust (2021) found that employees in organisations with a culture of experimentation were 37% more likely to engage with AI-driven analytics tools proactively. Similarly, a 2023 McKinsey survey reported that firms promoting "digital curiosity" saw significantly higher self-reported willingness to test generative AI models than those with more rigid, hierarchical norms (Chui et al., 2023).

Conversely, control-oriented or risk-averse cultures often suppress open discussion about AI. Workers in such settings fear being judged as "lazy" or "replaceable" if they admit to using automation tools (Fountaine et al., 2019). This echoes the early days of computerisation, when employees concealed spreadsheet automation to preserve perceived expertise. The current "AI stigma" mirrors that same defensive behaviour, especially in traditional industries where managerial cultures prioritise control over innovation. An equally significant factor is psychological safety

- the collective belief that one can voice ideas or admit mistakes without fear of reprisal (Edmondson, 2018). In cultures where psychological safety is low, employees hesitate to discuss AI usage, particularly in knowledge-intensive sectors where professional identity is tied to intellectual performance. On the other hand, organisations that normalise experimentation and treat AI as an augmenting rather than replacing technology report greater cross-departmental collaboration (McKinsey & Company, 2022). Practical case studies from large consulting and technology firms (e.g., PwC; IBM) demonstrate that framing AI as a human-centred, augmentative capability - emphasising collaboration (e.g., "human-AI collaboration" or "AI as a talent multiplier") - helps increase voluntary adoption and trust among employees. Empirical reports and practitioner playbooks from these organisations recommend messaging, governance and skills programmes that position AI as an enabler of better decisions rather than a replacement for human judgement (PwC, 2025). This cultural signalling reassures employees that AI use is a mark of competence rather than a confession of inadequacy. Furthermore, transparent communication about data governance and ethical safeguards fosters a climate of trust, making employees more inclined to share AI-based insights with peers (Raisch & Krakowski, 2021). In summary, organisational culture functions as both the accelerator and the brake of AI adoption. Cultures that reward learning, collaboration, and openness create a fertile ground for AI-driven transformation. Those anchored in rigid hierarchies and implicit fear of replacement, however, generate "digital silence," where AI tools are used secretly, without shared learning. There is a sizable minority of workers in every generational cohort who are afraid of technology even without understanding it, so-called "technophobes", and it is a recursive category, since such workers appear with every technological advancement, but this time it seems to be more serious than when the computerisation was implemented (McClure, 2017). Therefore, cultivating a psychologically safe, trust-based, and learning-oriented culture is not merely an HR recommendation - it is the precondition for responsible and innovative AI integration.

6. PSYCHOLOGICAL AND MANAGERIAL MECHANISMS BEHIND THE CONCEALMENT OF AI USE

The concealment of AI use in contemporary workplaces has become a subtle yet widespread phenomenon. Employees increasingly rely on generative and analytical AI systems, but often hide this reliance from supervisors or peers. The roots of such behaviour are both psychological and managerial, stemming from deep-seated cultural norms and structural incentives within organisations. From a psychological standpoint, the first mechanism is impression management - the human tendency to control how one is perceived by others. When workers start to believe that their AI usage signals laziness or incompetence, they downplay or hide their interaction with these tools (Huang & Rust, 2018). This tendency is especially strong in knowledge-intensive sectors, where professional identity is tied to intellectual effort rather than digital facilitation. Employees thus fear that revealing AI assistance could devalue their expertise or originality (Raisch & Krakowski, 2021). The discrepancy between the admitted use and the real use of AI is quite visible: 66% of individuals report that they do not use AI or use it only minimally when they are asked generally, but 43% of these respondents acknowledge that they regularly or sometimes use business products with AI if the question is more specific about the tasks (Ransbotham et al., 2022). The second psychological mechanism is AI anxiety, a composite of fear of replacement and loss of autonomy (Fountaine et al., 2019). Even when employees recognise AI's utility, they may experience "technological shame" - a reluctance to admit reliance on algorithms perceived as competitors rather than collaborators. Such anxiety suppresses open dialogue and creates a paradoxical situation where AI is both ubiquitous and invisible in everyday workflows (McKinsey & Company, 2022). Managerial mechanisms reinforce this silence.

Traditional performance appraisal systems tend to reward visible individual effort rather than outcome efficiency (Edmondson, 2018). As a result, workers are disincentivised from admitting that AI contributed to their productivity. Moreover, a lack of formal policies on acceptable AI use leaves employees uncertain about managerial attitudes. Several surveys indicate that a significant share of employees are using generative AI tools without informing their managers, often due to uncertainty about company policies or fear of disciplinary action. This ambiguity creates what Edmondson (2018) terms a "psychologically unsafe climate," where silence is perceived as the safer route. Furthermore, managerial opacity - the absence of transparent communication about AI's strategic role - deepens mistrust. When leaders fail to articulate whether AI is intended to augment or replace human labour, employees interpret every automation tool as a potential threat. In such contexts, concealment becomes a defensive survival strategy, protecting both job security and self-image (Chui et al., 2023). Ultimately, the concealment of AI use illustrates a broader organisational paradox: firms invest heavily in digital transformation yet maintain managerial cultures that penalise openness about it. Overcoming this requires reconfiguring both managerial systems and collective mindsets to treat AI as a legitimate extension of human capability rather than a sign of weakness or redundancy.

7. HOW GENERATION Z CHALLENGES AND RESHAPES TRADITIONAL MANAGERIAL EXPECTATIONS IN THE CONTEXT OF AI ADOPTION

Generation Z, born roughly between 1997 and 2012, is the first truly "AI-native" workforce cohort - raised alongside algorithms, recommendation engines, and machine learning interfaces. Their entry into the labour market has disrupted long-held managerial expectations about hierarchy, communication, and technological adaptation. Unlike previous generations that learnt to trust technology, Gen Z workers expect seamless integration of digital tools, including AI, as a normal part of everyday work (Francis & Hoefel, 2018). This generational disposition challenges traditional management in three key ways. First, Gen Z demands participatory decision-making and horizontal collaboration, rather than top-down control. Research by Deloitte (2025) indicates that Gen Z employees prefer managers who act as "mentors" rather than "supervisors." In the context of AI, this means they expect transparency and co-creation when new tools are introduced, not unilateral mandates. When leaders fail to include them in AI discussions, young employees often respond with disengagement or silent resistance (Chui et al., 2023). Their expectations are the following: when a human works together with AI, the human will be less constrained computationally, and the human-machine teams should make better decisions than humans alone (Page & Kallapur, 2025), not meaning less dull work and no additional tasks.

Second, Gen Z redefines what counts as professional competence. For them, using AI is not a shortcut but a demonstration of digital fluency and critical thinking. PwC's "Global Workforce Hopes & Fears 2024" states that among employees who use generative-AI daily, 82% expect it to make their work more efficient. While no exact generational ratio is publicly confirmed, large-scale surveys from PwC indicate younger workers (such as Gen Z) are significantly more likely to engage with generative AI tools than older cohorts, especially Baby Boomers (PwC, 2024). This contrasts with older managerial norms that equate visible effort with value. Consequently, friction arises when supervisors interpret AI reliance as a lack of diligence rather than as an intelligent optimisation strategy (Fountaine et al., 2019). Third, Gen Z's value system is closely tied to authenticity, ethics, and purpose, extending into their perception of AI. They are more likely to question opaque data practices, algorithmic bias, and the social impact of automation (Deloitte, 2025). Traditional managers, accustomed to technocratic adoption, are often unprepared for such ethical scrutiny from subordinates. This shift compels leaders to become educators and ethical

stewards, not just implementers of efficiency. Practically, this generational tension has prompted some companies to establish "reverse mentoring" programmes, where Gen Z employees advise senior managers on AI usage and digital trends (McKinsey & Company, 2022). These initiatives symbolise a cultural inversion of expertise - authority no longer flows strictly from experience but also from digital intuition. As Raisch and Krakowski (2021) note, successful AI transformation requires precisely this type of reciprocal learning, where managers unlearn outdated notions of control and embrace distributed intelligence. In sum, Generation Z is not merely adapting to AI-driven change - they are redefining its cultural logic. Their expectations for openness, ethics, and collaboration are forcing organisations to evolve from hierarchical command structures into networked, learning ecosystems. The question is no longer whether Gen Z will adapt to managerial norms, but whether management itself can adapt to them. If the management fails now, it would be even harder with Gen Alpha, which is said to be even less flexible and more demanding than the previous generation (Crumley, 2025).

8. FUTURE RESEARCH DIRECTIONS

While no primary data (such as surveys or interviews) was collected in this phase, future research may include quantitative methods to test the hypotheses and themes presented here.

9. CONCLUSION

The convergence of technological, psychological, and generational forces ("the perfect storm" in HR management in the last decade) has created a profound shift in how organisations approach AI adoption. The studies reviewed here suggest that the success or failure of AI integration depends less on the technology itself and more on the human environment surrounding it. Organisational culture, psychological safety, and generational expectations jointly define whether AI becomes a shared instrument of progress or a hidden, mistrusted companion. A culture that values transparency, learning, and experimentation encourages employees to use and discuss AI tools openly, transforming the technology into a collaborative asset rather than a secret advantage. Conversely, when fear, hierarchy, or managerial opacity dominates, employees resort to concealment - protecting their self-image and job security instead of collective innovation. This concealment phenomenon underscores the persistence of industrial-era control logics within post-digital organisations. At the same time, Generation Z's arrival is accelerating this cultural reckoning. Their comfort with AI, demand for authenticity, and ethical sensitivity directly challenge managerial traditions built on surveillance and procedural authority. Gen Z expects a workplace where AI enhances human creativity, not one where it replaces or monitors it. Their influence is pushing organisations toward horizontal learning models, where expertise flows both upward and downward across generations. The implication is clear: sustainable AI transformation requires a human-centric recalibration of management itself. In its efforts to avoid the perception of AI adoption as a mutation of "working smarter, not harder" mutates into an implicit expectation of "working faster, for longer", managers must change their own managerial approaches in order to assure the frankness of their employees and their trust concerning the real AI usage. As we know from history, great painters never worked all alone on all their masterpieces - they usually made sketches, and their pupils finished the paintings with the master's colours and mannerism - it did not make the paintings less valuable or "less theirs". In the aviation industry, pilots did not become obsolete with the arrival of automatic pilots, nor did their salaries suffer a downgrade. The same should apply in the other areas of business. Efficiency gains thus should be redistributed to benefit employees, and not reinvested into more demanding output targets, as we mentioned in the beginning. Finally, the answer to the question "Should AI be an autopilot, a copilot, a servant, a competitor, or

a surveillance tool?“ should therefore be „a copilot“, and there is no wonder that one of the most advanced AI tools bears that very name. Therefore, leaders must evolve from supervisors into facilitators, HR must become the guardian of psychological safety, and employees - especially younger ones - must be empowered as co-creators in the digital ecosystem with no fear that they would be fired or burdened with more tasks. When these elements align, AI adoption ceases to be a threat to human value and becomes, instead, a reflection of it. The “perfect HRM storm” thus becomes an opportunity: a chance to rebuild organisational life on trust, openness, and shared intelligence rather than fear and control. It is therefore useful to promote the perception of AI tools as a “copilot” that allows the worker to focus on the creative and strategic components of the job, while routine or data-intensive tasks are taken over by the AI, alleviating the fear of being redundant and expelled from the workplace.

Acknowledgements

The authors are grateful for the support within the project of the Department for General Disciplines in Technology, Faculty of Technical Sciences Novi Sad entitled “Socio-technological aspects of improving the teaching process in the English language in fundamental disciplines”.

References

Acemoglu, D. (2025). The simple macroeconomics of AI, *Economic Policy*, *40*(121), 13–58, https://doi.org/10.1093/epolic/eiae042

Burleigh, E. (2025). *Billionaire bosses like Jeff Bezos and Reid Hoffman denounce work-life balance—and some think working nonstop is key to success*, https://fortune.com/2025/10/15/billionaire-ceo-jeff-bezos-reid-hoffman-work-life-balance-careers-business-boundaries-success-entrepreneurs-38-hour-workweek/

Burrell, J. (2016). How the machine ‘thinks’: Understanding opacity in machine learning algorithms. *Big Data & Society*, *3*(1). https://doi.org/10.1177/2053951715622512

Chui, M., Hazan, E., Roberts, R., Singla, A., Smaje, K., Sukharevsky, A., Yee, L., & Zemmel, R. (2023). *The Economic Potential of Generative AI: The Next Productivity Frontier*, McKinsey & Company.

Chung, C. H., & Schiff, D. S. (2025). AI and the Social Contract. *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, *8*(1), 615-627. https://doi.org/10.1609/aies.v8i1.36575

Crist, C. (2025). 42% of Gen Z workers say they’re turning to blue-collar roles for security, HR Dive, www.hrdive.com/news/gen-z-workers-say-theyre-turning-to-blue-collar-roles-for-security/749115

Crumley, B. (2025). Generation Alpha May Find the Workplace Even Tougher Than Gen-Z Does, Inc.com, *www.inc.com/bruce-crumley/generation-alpha-may-find-the-workplace-even-tougher-than-gen-z-does/91248959*

Davenport, T. H., & Mittal, N. (2023). *All-in on AI: How smart companies win big with artificial intelligence.* Harvard Business Review Press.

Deloitte. (2025). *2025 Gen Z and Millennial Survey: Growth and the pursuit of money, meaning, and well-being*

Edmondson, A. (2018). *The Fearless Organisation: Creating Psychological Safety in the Workplace for Learning, Innovation, and Growth*. Hoboken, NJ: John Wiley & Sons.

Farrell, H., Gopnik, A., Shalizi, C., & Evans, J. (2025). Large AI models are cultural and social technologies. *Science, 387*(6739), 1153-1156. https://doi.org/10.1126/science.adt9819

Fountaine, T., McCarthy, B., & Saleh, T. (2019). Building the AI-powered organisation. *Harvard Business Review, 97*(4), 62–73.

Francis, T., & Hoefel, F. (2018). True Gen: Generation Z and its implications for companies. *McKinsey & Company.*

Fuchs, O., Lorenz, E., & Fuchs, L. (2024). Generational differences in attitudes towards work and career: A systematic literature review on the preferences of generations X, Y, and Z. *International Journal of Innovative Research and Advanced Studies*, *11*(7), 54-71.

Grasiaswaty, N. (2025). Algorithmic human resource management: A brief introduction and managerial toolkit. *EWOP in Practice*. *19*(1), 61–76. https://doi.org/10.21825/ewopinpractice.94698.

Haenlein, M., Kaplan, A., Tan, C. W., & Zhang, P. (2019). Artificial intelligence (AI) and management analytics. *Journal of Management Analytics*, *6*(4), 341–343. https://doi.org/10.1080/23270012.2019.1699876

Huang, M.-H., & Rust, R. T. (2018). Artificial Intelligence in Service. *Journal of Service Research*, *21*(2), 155-172. https://doi.org/10.1177/1094670517752459

Huang, M.-H., & Rust, R. T. (2021). Engaged to a robot? The role of AI in service. *Journal of Service Research, 24*(1), 30–41. https://doi.org/10.1177/1094670520902266

Kellogg, K. C., Valentine, M. A., & Christin, A. (2020). Algorithms at work: The new contested terrain of control. *Academy of Management Annals, 14*(1), 366–410. https://doi.org/10.5465/annals.2018.0174

Kraus, S., Palmer, C., Kailer, N., Kallinger, F. L., & Spitzer, J. (2019). Digital entrepreneurship: A research agenda on new business models for the twenty-first century. *International Journal of Entrepreneurial Behavior & Research*, 25(2), 353–375. https://doi.org/10.1108/ijebr-06-2018-0425

Lee, M. K., Kusbit, D., Metsky, E., & Dabbish, L. (2015). Working with machines: The impact of algorithmic and data-driven management on human workers. *Proceedings of the 33rd ACM Conference on Human Factors in Computer* Systems (CHI '15), Association for Computing Machinery, New York, NY, USA, 1603 – 1612, https://doi.org/10.1145/2702123.2702548

Mateescu, A., & Nguyen, A. (2019). *Explainer: Algorithmic management in the workplace.* Data & Society Research Institute.

McAfee, A., & Brynjolfsson, E. (2017). *Machine, platform, crowd: Harnessing our digital future.* WW Norton & Company.

McClure, P. K. (2017). "You're Fired," Says the Robot: The Rise of Automation in the Workplace, Technophobes, and Fears of Unemployment. *Social Science Computer Review*, *36*(2), 139-156. https://doi.org/10.1177/0894439317698637

McKinsey & Company. (2022). *The State of AI in 2022.* https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai-in-2022-and-a-half-decade-in-review

Möhlmann, M., & Zalmanson, L. (2017). Hands on the wheel: Navigating algorithmic management and Uber drivers' autonomy. 38th International Conference on Information Systems (ICIS 2017)

Page, S. E., & Kallapur, A. (2025). Replace, augment, disrupt: AI & organizational decision-making. *Journal of Organization Design*. https://doi.org/10.1007/s41469-025-00194-4

PwC. (2024). PwC 2024 Global Workforce Hopes & Fears Survey: Workers embrace AI and prioritise skills growth amid rising workloads and an accelerating pace of change. PricewaterhouseCoopers Global Report.

PwC. (2025). *SXSW 2025 key insights.* PricewaterhouseCoopers.

Rahwan, I., Cebrian, M., Obradovich, N., Bongard, J., Bonnefon, J. F., Breazeal, C., Crandall, J. W., Christakis, N. A., Couzin, I. D., Jackson, M. O., Jennings, N. R., Kamar, E., Kloumann, I. M., Larochelle, H., Lazer, D., McElreath, R., Mislove, A., Parkes, D. C., Pentland, A. S., 1, Roberts, M. E., Shariff, A., Tenenbaum, J. B., & Wellman, M. (2019). Machine behaviour. *Nature, 568*, 477–486. https://doi.org/10.1038/s41586-019-1138-y

Raisch, S., & Krakowski, S. (2021). Artificial intelligence and management: The automation–augmentation paradox. *Academy of Management Review, 46*(1), 192–210. https://doi.org/10.5465/amr.2018.0072

Ransbotham, S., Kiron, D., Candelon, F., Khodabandeh, S., & Chu, M. (2022). Achieving Individual — and Organisational — Value With AI, MIT Sloan Management Review, https://sloanreview.mit.edu/projects/achieving-individual-and-organizational-value-with-ai/

Resume Builder. (2025). *4 in 10 Gen Z College Grads Are Turning To Blue-Collar Work for Job Security*, www.resumebuilder.com/4-in-10-gen-z-college-grads-are-turning-to-blue-collar-work-for-job-security

Rosenblat, A. (2018). *Uberland: How algorithms are rewriting the rules of work.* University of California Press. https://doi.org/10.1525/9780520970632

Schein, E. H., & Schein, P. A. (2017). *Organisational Culture and Leadership, 5th edition*. John Wiley & Sons.

Schroth, H. (2019). Are you ready for Gen Z in the workplace? *California Management Review, 61*(3), 5–18. https://doi.org/10.1177/0008125619841006

Susskind, R., & Susskind, D. (2015). The Future of the Professions: How Technology Will Transform the Work of Human Experts. Oxford University Press, https://doi.org/10.1093/oso/9780198713395.001.0001

Balkans JETSS (2025) 2: 114-122
DOI: https://doi.org/10.31410/Balkans.JETSS.2025.8.2.114-122

Original Scientific Article

Integrating DEI Into ESG Strategies: A Social Perspective in Higher Education Institutions

Goran Pavlović[1]
Olga Pavlović[2]

Received: October 14, 2025 / Revised: November 20, 2025 /
Accepted: December 15, 2025 / Published: December 30, 2025

Abstract: *This study examines the role of Diversity, Equity, and Inclusion (DEI) as a key operational component of the social dimension within Environmental, Social, and Governance (ESG) frameworks in Higher Education Institutions (HEIs). Drawing on contemporary ESG and DEI theory, the paper conceptualizes DEI as a measurable and practice-oriented mechanism through which social sustainability can be strengthened in academic environments. The empirical research was conducted in the first quarter of 2025 and included undergraduate, master's, and doctoral students from two HEIs in Serbia. Data were collected using a structured questionnaire based on a validated multicultural competence framework and measured on a five-point Likert scale. Quantitative analysis involved descriptive statistics, independent samples t-tests to examine gender differences, and one-way ANOVA to assess differences across levels of study. The findings indicate a very high overall perception of DEI practices, particularly regarding faculty support, inclusive academic climate, and students' openness to working with individuals from diverse backgrounds. Statistically significant gender differences were observed in several DEI dimensions, with female students reporting higher awareness and perceived cultural competence, while no significant differences were found across levels of study. The study provides practical insights for integrating DEI into ESG strategies in HEIs, supporting institutional social responsibility, inclusive governance, and long-term social sustainability.*

Keywords: *Diversity, Equity, Inclusion, HEI, ESG, Social sustainability*

JEL Classification: I23 · I24

✉ goran.pavlovic@metropolitan.ac.rs
[1] Belgrade Metropolitan University, Tadeuša Košćuška 63, 11158 Belgrade, Serbia
[2] Belgrade Metropolitan University, Tadeuša Košćuška 63, 11158 Belgrade, Serbia

1. INTRODUCTION

In the contemporary higher education system, the concepts of sustainability and social responsibility are gaining increasing importance. Integrating the DEI principles into the core operational strategies of HEIs has become essential for achieving a socially responsible university system. DEI components contribute to the creation of a fairer academic environment that values diversity and promotes equal access to educational and professional opportunities. In this context, the ESG approach in higher education goes beyond the traditional understanding of sustainability, as it integrates environmental, social, and governance dimensions. The primary mission of HEIs is the creation and dissemination of knowledge; however, they are also tasked with equipping future generations with the necessary skills and competencies to effectively address sustainability challenges (Alenezi & Alanazi, 2024). Today, sustainability is increasingly linked to social issues and challenges. For this reason, many HEIs have incorporated DEI principles into their institutional missions, demonstrating a strong commitment to fostering an inclusive academic environment (Fagun, 2025).

Within the DEI framework, diversity refers to the recognition of physical and social characteristics that distinguish individuals or groups, emphasizing these differences as sources of strength. Equity, as a key component of the DEI approach, ensures equal opportunities and access regardless of background by removing barriers that hinder participation and creating fair conditions for all. Inclusion represents the active appreciation of diverse perspectives and the engagement of individuals from all social backgrounds, fostering a sense of belonging and recognition of each community member for who they truly are (Yanikoglu, 2025). The various forms of diversity that must be integrated into HEIs' strategic decisions relate to socio-demographic characteristics such as gender, age, ethnicity, religion, sexual orientation, beliefs, lifestyle, nationality, physical characteristics, and many others (Bele & Hebalkar, 2023; Lumadi, 2008).

This research aims to examine students' perceptions of the implementation of DEI principles within their higher education institutions, with particular emphasis on inclusivity, equal opportunities, and campus diversity. Based on this objective, the following research question has been formulated:

Q1: How do students perceive the implementation of diversity, equity, and inclusion (DEI) principles within their higher education institutions, and in what ways do these practices influence their academic experiences and institutional engagement?

2. LITERATURE REVIEW

In the last few years, the concept of ESG has become more important in higher education. Many universities try to connect their main goals with sustainable development and social responsibility. At the beginning, ESG was created to follow and evaluate companies, mostly in the financial and business fields. Today, it is used in a much wider context to observe how institutions behave and how responsible they are toward society (Li et al., 2021).

The environmental and governance parts of ESG are already well structured and measurable, but the social part (S) is still less researched and not clearly defined (Boubaker et al., 2018). This social pillar includes equality, human rights, diversity, and the relationship between institutions and the community. Still, these values are difficult to measure precisely, since they depend on people's perception and cultural background.

The idea of DEI has changed much in the last decade. It shows the social, organizational, and cultural changes that happen in workplaces today. In the study of Park et al. (2025), DEI was first used as a way to correct unfair situations in management and organization. The goal was to reduce inequality in hiring and promotion. Later, the DEI idea became wider and included also topics of organizational culture, leader responsibility, and justice inside institutions. Verma (2024) explains that diversity means differences between people in the work environment, like gender, ethnicity, age, or social background. Equity means fairness in systems and rules, knowing that not all people start from the same position. Inclusion means active participation and a feeling of belonging, when every person is respected, listened and supported. All these together make DEI not only a formal rule but a real practice that helps to create fairness and real participation in organizations.

Because of that, many researchers include the concept of DEI to explain the social side of ESG in a more practical and visible way. DEI helps to turn social ideas into real institutional actions and policies that can be compared and evaluated (Roberson, 2019; Shore et al., 2011).

Recent studies see DEI also as a moral and practical need. Wang et al. (2024) found that DEI has a strong effect on work culture and on the well-being of workers, especially through anti-racism programs, trainings, and development activities. They write that DEI works best when it is a long-term plan with real support from leaders, not only one short event. In the same way, Hattery et al. (2022) show that diversity in teams can bring more creativity, innovation, and better decision-making. But they also say that fairness and inclusion need active management, otherwise it becomes only formal or surface-level. The authors see DEI not only as an ethical value but as a way to support learning and to make organizations more strong. In general, these studies show that DEI is now a basic part of sustainable and responsible institutions, because it joins fairness, participation, and representation on all levels.

Several studies show that organizations with developed DEI practices achieve better ESG results. Gidage's (2025) study found that companies that promote diversity and inclusion receive higher ESG scores and gain more trust from employees and the public. The same authors also point out that DEI-based management improves adaptability and long-term resilience. These results show that DEI is not only about ethics, but also about the sustainability and growth of institutions.

In universities, ESG is becoming a practical framework for understanding how they contribute to society and equality inside their structures (Alenezi & Alanazi, 2024). However, environmental activities are still more visible, while the social part is less developed (Dutta et al., 2025). Including DEI principles in ESG strategies allows universities to build inclusive policies, ensure equal opportunities, and support students and staff through a positive and safe environment.

A good example is the study of Smith et al. (2017), which examined how students experience diversity and cultural competence at universities. More than 90% of the students said that teachers and staff were open and accepting of everyone. Most of them also learned about different cultures and were ready to work with people of any background. This shows that an inclusive environment supports the development of intercultural and professional skills. The same authors also expanded the idea of diversity beyond ethnicity, adding gender, religion, disability, and socioeconomic background. These factors help to understand how universities can promote fairness, equality, and cultural understanding. From this point of view, DEI is the most practical and measurable way to develop the social part of ESG in higher education.

Newer studies confirm this link. According to Yanıkoglu (2025), the language and communication used at universities strongly affect the level of inclusion. Authors also stress that wider access to education requires fairness and accessibility. These results show that DEI should be seen as a permanent part of institutional culture and long-term sustainability strategy, not just as a one-time initiative.

Altogether, previous studies (Alenezi & Alanazi, 2024; Li et al., 2021; Roberson, 2019; Smith et al., 2017) clearly show that ESG becomes complete only when it includes DEI principles. With DEI, institutions can create inclusive environments, support social sustainability, and strengthen social cohesion in the academic community.

3. METHODOLOGY

This paper aims to examine the role of DEI as key indicators of the social dimension in ESG frameworks within HEIs. The research aims to assess students' perceptions of DEI practices in their institutions, focusing on inclusivity, equal opportunities, and campus diversity. The study objectives were to identify students' views about diversity in the department and university, perceptions of faculty's acceptance and accommodation of diversity, and knowledge and development of skills related to cultural competence. Quantitative data were collected and analyzed to determine if objectives were met. The sample included students of undergraduate studies, and postgraduate students (Master's and doctoral students) at the Belgrade Metropolitan University and the University of Kragujevac. Data collection was carried out during the first quarter of 2025.

The multicultural competence questionnaire was designed according to the study of Smith et al. (2017), in order to measure students' perceptions of diversity, inclusion, and multicultural competence during their studies. Perceptions were measured on a five-point Likert scale, ranging from 1 (strongly disagree) to 5 (strongly agree). The questionnaire included statements related to attitudes of faculty, staff, and students toward diversity. Students were asked if faculty and staff accept all individuals regardless of differences, include multicultural topics in the study program, and use neutral and inclusive language. Questions also examined whether minority students were treated equally, whether faculty were open and supportive to all students, and whether they avoided stereotypes based on age or gender. Additional items asked if faculty made adjustments for religious reasons (for example, changing assignment deadlines) and if the faculty provided information in accessible formats, such as video materials, to meet the needs of all students. Students were also asked if their peers were open and supportive toward everyone, including minority groups. Finally, the questionnaire measured students' personal development: whether they developed multicultural competencies, increased knowledge about people from different backgrounds, and felt open to cooperation with people from different cultures. Demographic data included gender, level of study, and year of study (for undergraduate students only).

The sample is characterized by a higher representation of male respondents (69.64%), while females make up 30.36% of the total. Most participants are undergraduate students (60.71%), followed by PhD students (28.58%) and master's students (10.71%). This distribution shows that the sample includes students from different academic levels, providing a balanced overview of perceptions among individuals at various stages of their studies. The predominance of undergraduate students is expected, since universities generally have a larger number of students enrolled in undergraduate programs compared to master's and doctoral studies. Therefore, the difference in representation by study level can be considered natural and consistent with the overall student population structure. This structure allows comparison of attitudes toward multicultural competence and inclusion across different study levels, which contributes to the validity and representativeness of the research findings.

Table 1. Sample characteristics

	Frequency (N=56)	
	N	%
Gender		
Female	17	30,36%
Male	39	69,64%
Level of study		
Undergraduate studies	34	60,71%
Master's students	6	10,71%
PhD students	16	28,58

Source: Authors

4 RESULTS AND DISCUSSION

At the beginning of the analysis, a descriptive statistical analysis was applied in order to determine the average value (M) and standard deviation (SD) of the items used to measure the variables of the model. The results are shown in Table 2.

Table 2. Results of descriptive statistics

Items	M	SD
Faculty and staff in the department were accepting of all individuals (regardless of diversity).	4,8214	,38646
Multicultural topics were addressed (within my curriculum).	4,1964	,79589
Faculty were open to learning about different cultures.	4,3929	,86715
Faculty used neutral language indicating acceptance (of varied sexual orientation).	4,2143	,92862
Minority students were not singled out (to represent the options of entire minority population).	4,4821	,97218
Faculty in the major were open and supportive of all students (including minority students).	4,7143	,59435
Faculty avoided stereotyping by age.	4,7500	,51346
Faculty used gender neutral language.	4,0536	,99854
Faculty accommodated students for religious accommodations.	3,8571	1,48236
Faculty presented information in accessible formats (such as closed captioned videos) to meet all students' needs.	4,3036	1,23465
Students in the major were open and supportive of all students (including minority students).	4,2500	,91949
I have developed the skills to be a culturally competent health professional.	4,1250	1,12916
I have increased my knowledge of people with diverse backgrounds.	4,2679	1,03557
I am open to working with people with diverse backgrounds.	4,9821	,13363

Source: Authors

On a five-point scale, the average value of DEI findings is 4.38, which can be characterized as very high, and which leads to the conclusion that there are DEI practices and strategies in HEIs. Among all measured items, the statement "*I am open to working with people with diverse backgrounds*" obtained the highest mean value (M = 4.98). This result indicates a very high level of agreement among respondents and reflects students' strong openness and willingness to collaborate with individuals from different cultural, ethnic, and social backgrounds. Such a result suggests that the participants possess well-developed intercultural awareness and acceptance, which are essential components of DEI-related competencies. It also implies that HEIs have likely

succeeded in fostering an academic culture that values diversity and prepares students for professional environments characterized by inclusivity and global collaboration. At the same time, this item had the lowest standard deviation (SD = 0.13), which, as such, indicates a high degree of homogeneity of students' attitudes.

The second-highest mean value was recorded for the statement *"Faculty and staff in the department were accepting of all individuals (regardless of diversity)"* (M = 4.82), suggesting that students largely perceive their academic environment as open and welcoming to diversity. This perception reflects a generally inclusive institutional culture, where faculty and staff demonstrate respect and support for all members of the academic community, regardless of their background. Closely following is the item *"Faculty in the major were open and supportive of all students (including minority students)"* (M = 4.71), which further reinforces the idea that teaching staff play a significant role in promoting inclusivity and ensuring equal participation of all students. The item *"Faculty avoided stereotyping by age"* (M = 4.75) also indicates a strong awareness among educators regarding potential biases, suggesting that age-related stereotypes are effectively minimized within the learning environment.

Moderately high means were found for *"Minority students were not singled out to represent the opinions of entire minority populations"* (M = 4.48) and *"Faculty were open to learning about different cultures"* (M = 4.39). These results suggest that faculty members are generally sensitive to issues of representation and are willing to expand their cultural understanding, contributing to a respectful and equitable classroom climate. The item *"Faculty presented information in accessible formats (such as closed-captioned videos) to meet all students' needs"* (M = 4.30) also scored relatively high, reflecting institutional efforts to accommodate diverse learning needs and promote accessibility. Items related to curriculum content and communication practices, such as *"Multicultural topics were addressed within my curriculum"* (M = 4.19) and *"Faculty used neutral language indicating acceptance of varied sexual orientations"* (M = 4.21), showed positive but slightly lower averages, indicating room for further integration of DEI topics into teaching materials and language practices. Meanwhile, *"Students in the major were open and supportive of all students (including minority students)"* (M = 4.25) suggests a positive peer culture that aligns with institutional DEI efforts.

On the other hand, *"I have increased my knowledge of people with diverse backgrounds"* (M = 4.27) and *"I have developed the skills to be a culturally competent health professional"* (M = 4.13) demonstrate that students recognize personal growth in cultural competence but may still require more experiential learning opportunities to strengthen these skills. The lowest mean was observed for *"Faculty accommodated students for religious accommodations"* (M = 3.86), indicating that while general inclusivity is well established, religious accommodation remains an area where further institutional attention and support are needed. In order to determine the differences in the attitudes of male and female students, a t-test was applied, the results of which are shown in Table 3.

The t-test results revealed statistically significant gender differences across several DEI-related items, with women students reporting higher mean scores than their male counterparts. Female respondents perceived that multicultural topics were more frequently addressed within their curriculum and that faculty demonstrated greater openness to learning about different cultures. This finding may suggest that women are generally more attuned to diversity-related content and possibly more responsive to inclusive pedagogical practices. It also aligns with previous research indicating that women tend to show stronger sensitivity toward multiculturalism and social equity

issues in educational settings. Furthermore, women students rated higher agreement with the statements that faculty used neutral language, indicating acceptance of varied sexual orientations and that minority students were not singled out to represent entire minority populations. These perceptions reflect a stronger awareness of linguistic inclusivity and representational fairness among female students, which may stem from a heightened sensitivity to social identity and discrimination dynamics. The finding highlights that female students might experience or notice inclusive practices more readily, possibly due to their own experiences. The gender difference was also evident in the items "Faculty accommodated students for religious accommodations" and "I have developed the skills to be a culturally competent health professional." Higher female scores in these areas suggest a stronger perception of institutional support for diverse religious needs and greater self-assessed competence in intercultural understanding. This may indicate that women students engage more deeply with DEI-related learning opportunities, particularly those fostering empathy, reflection, and communication across cultural boundaries. Overall, these findings imply that women students not only perceive a more inclusive and equitable academic environment but also internalize DEI principles more effectively.

Table 3. Results of the t-test

Items	Mean difference	p value
Faculty and staff in the department were accepting of all individuals (regardless of diversity).	-0.87	0.441
Multicultural topics were addressed (within my curriculum).	-0.47	0.038*
Faculty were open to learning about different cultures.	-0.53	0.046*
Faculty used neutral language indicating acceptance (of varied sexual orientation).	-0.45	0.006*
Minority students were not singled out (to represent the options of entire minority population).	0.26	0.034*
Faculty in the major were open and supportive of all students (including minority students).	-0.24	0.439
Faculty avoided stereotyping by age.	-0.19	0.077
Faculty used gender neutral language.	-0.85	0.206
Faculty accommodated students for religious accommodations.	-0.62	0.003*
Faculty presented information in accessible formats (such as closed captioned videos) to meet all students' needs.	-0.07	0.147
Students in the major were open and supportive of all students (including minority students).	-0.65	0.845
I have developed the skills to be a culturally competent health professional.	-0.15	0.013*
I have increased my knowledge of people with diverse backgrounds.	-0.12	0.547
I am open to working with people with diverse backgrounds.	-0.05	0.592

Note: * - Results are significant at 0,05 level

Source: Authors

The results of ANOVA test revealed no statistically significant differences among undergraduate, master's, and doctoral students regarding their perceptions of DEI-related practices within their institutions. This finding suggests a relatively consistent level of awareness and experience with diversity, equity, and inclusion across all levels of study. Such uniformity may indicate that DEI principles are embedded institutionally rather than being limited to specific academic levels or programs. In other words, students appear to encounter similar inclusive practices, faculty behaviors, and institutional policies throughout their academic progression. The absence of variation across study levels also implies that HEIs have established a stable and coherent approach to DEI implementation. This consistency may result from university-wide initiatives, such as faculty training, standardized curricula, or campus-wide policies promoting inclusivity and equity. It

is also possible that DEI topics are introduced early in students' educational experience, allowing for the formation of stable perceptions that persist through advanced stages of study. However, while the lack of significant differences is generally positive, it also points to potential limitations in the differentiation or deepening of DEI engagement at higher academic levels. Ideally, graduate and doctoral programs should provide more advanced and context-specific opportunities for critical reflection on diversity and social responsibility. Therefore, future institutional efforts might consider tailoring DEI education and experiences to align with the increasing academic and professional maturity of students, ensuring that inclusivity remains not only consistent but also progressively enriching throughout higher education.

5. CONCLUSION

Students perceive the implementation of DEI principles within their HEIs as highly positive, indicating that these institutions foster inclusive, equitable, and supportive academic environments. They recognize that faculty and staff demonstrate openness, respect, and a commitment to accommodating diverse cultural, social, and individual needs, which contributes to a sense of belonging and fairness across the campus. DEI practices are perceived to enhance students' academic experiences by promoting intercultural understanding, reducing stereotyping, and creating opportunities for collaboration with peers from diverse backgrounds. These practices also influence institutional engagement, as students report feeling more valued and supported, which encourages active participation in academic and extracurricular activities. The findings suggest that students, particularly women, internalize DEI principles and develop cultural competence, which prepares them for professional environments characterized by inclusivity and global collaboration. Overall, the perception of effective DEI implementation strengthens both the quality of learning and the level of engagement within the HEI, fostering an academic culture that aligns with principles of social responsibility and equitable opportunity.

The limitations of this study primarily arise from the scope of data collection and the characteristics of the sample. First, the research was conducted at only two universities, which limits the extent to which the findings can be generalized to the broader population of higher education institutions. Differences in institutional policies, campus culture, and student demographics at other universities may result in variations in the implementation and perception of DEI principles that are not captured in this study. Second, the relatively small sample size further constrains the robustness of the statistical analyses, potentially affecting the reliability and precision of the results. While the descriptive statistics provide valuable insights into students' perceptions, the limited sample reduces the ability to detect subtle patterns or differences across subgroups, such as academic disciplines or specific demographic categories. Consequently, caution is warranted when interpreting these findings, and they should be considered indicative rather than definitive. Future research should aim to include a larger, more diverse sample across multiple HEIs to strengthen the generalizability of the results, allow for more sophisticated statistical analyses, and provide a deeper understanding of how DEI practices are perceived and experienced across different contexts in higher education.

ACKNOWLEDGMENT

This paper is the result of the work on Erasmus+ SHIFT - ESG Impact Index in Higher Education project, number 2023-1-ES01-KA220-HED-000152577.

References

Alenezi, M., & Alanazi, F. (2024). Integrating environmental social and governance values into higher education curriculum. *International Journal of Evaluation and Research in Education (IJERE), 13*(5), 3493. https://doi.org/10.11591/ijere.v13i5.29440

Bele, M., & Hebalkar, R. (2023). Significance of diversity, equity and inclusion (DEI) in higher educational institutions (HEI). *International Journal of Research Publication and Reviews, 4*(1), 1946–1954.

Boubaker, S., Cumming, D., & Nguyen, D. K. (2018). Introduction to the Research Handbook of Finance and Sustainability. *Research Handbook of Finance and Sustainability*. https://doi.org/10.4337/9781786432636.00006

Dutta, S., Dutta, R., Chatterjee, S., & Ray, S. (2025). Impact of DEI and Sustainability on Higher Educational Institutes in America. *Journal of International Commercial Law and Technology, 6*(1), 171-177. https://doi.org/10.61336/jiclt/25-01-12

Fagun, O. (2025). The Future of Diversity, Equity and Inclusion in Higher Education in United States. *International Journal of Research and Innovation in Social Science, IX*(IIIS), 1164-1175. https://doi.org/10.47772/ijriss.2025.903sedu0088

Gidage, M. (2025). Exploring the impact of diversity, equity and inclusion on ESG performance: evidence from Indian organizations. *Benchmarking: An International Journal*, 1-36. https://doi.org/10.1108/bij-11-2024-1009

Hattery, A. J., Smith, E., Magnuson, S., Monterrosa, A., Kafonek, K., Shaw, C., Mhonde, R. D., & Kanewske, L. C. (2022). Diversity, Equity, and Inclusion in Research Teams: The Good, The Bad, and The Ugly. *Race and Justice, 12*(3), 505-530. https://doi.org/10.1177/21533687221087373

Li, T.-T., Wang, K., Sueyoshi, T., & Wang, D. D. (2021). ESG: Research Progress and Future Prospects. *Sustainability, 13*(21), 11663. https://doi.org/10.3390/su132111663

Lumadi, M. W. (2008). Managing Diversity At Higher Education And Training Institutions: A Daunting Task. *Journal of Diversity Management (JDM), 3*(4), 1-10. https://doi.org/10.19030/jdm.v3i4.4996

Park, C. H., Park, S., & Kwon, B. (2025). Forty-five years of research on diversity, equity and inclusion in management. *Management Decision, 63*(13), 66-95. https://doi.org/10.1108/md-11-2023-2181

Roberson, Q. M. (2019). Diversity, equity, and inclusion in organizations: A review, critique, and future directions. *Annual Review of Organizational Psychology and Organizational Behavior*, 6, 69–88

Shore, L. M., Cleveland, J. N., & Sanchez, D. (2011). Inclusive workplaces: A review and model. *Human Resource Management Review*, 21(4), 311–326

Smith, T. M. E., Wessel, M. T., & Polacek, G. N. (2017). Perceptions of cultural competency and acceptance among college students: Implications for diversity awareness in higher education. *Journal of Cultural Diversity*, 24(3), 70–78.

Verma, A. (2024). Role of Diversity, Equity and Inclusion at Workplace. *International Journal of Business and Management Invention,* 13(1), 60-63.

Wang, M. L., Gomes, A., Rosa, M., Copeland, P., & Santana, V. J. (2024). A systematic review of diversity, equity, and inclusion and antiracism training studies: Findings and future directions. *Translational Behavioral Medicine, 14*(3), 156-171. https://doi.org/10.1093/tbm/ibad061

Yanıkoglu, Ö. (2025). Decoding Diversity, Equity, and Inclusion (DEI) in Higher Education: A Linguistic and Theoretical Exploration. *Educational Academic Research*(57), 146-159. https://doi.org/10.33418/education.1560565

Balkans JETSS (2025) 2: 123-137
DOI: https://doi.org/10.31410/Balkans.JETSS.2025.8.2.123-137

Original Scientific Article

Enhancing Hotel Management Through Predictive AI Models for Customer Lifetime Value (CLV)

Milena Nikolić[1]
Žarko Rađenović[2]
Marina Marjanović[3]

Received: October 10, 2025 / Revised: December 28, 2025 /
Accepted: December 29, 2025 / Published: December 30, 2025

Abstract: *This paper introduces a hybrid artificial intelligence (AI) framework for predicting Customer Lifetime Value (CLV) in hotel management. CLV represents the long-term financial contribution of guests and provides information for resource allocation, customer retention, and profitability. Traditional models rely on structured reservation records, often overlooking emotional insights in online reviews. To address this concern, the study combines structured booking attributes with unstructured guest reviews, using RoBERTa embeddings for textual data and XGBoost for numerical features. The proposed multimodal model achieves 89% accuracy, 87% precision, and 86% recall, outperforming single-source approaches utilizing bookings (63%) or guest reviews only (72%). SHAP-based interpretability reveals that review topics, including cleanliness, staff professionalism, and service quality, directly influence CLV, alongside structured features such as repeat bookings and special requests. The findings highlight the potential of predictive AI to enhance hotel management by identifying valuable customers early, supporting personalized services, and optimizing strategic decision-making.*

Keywords: *Hotel management, Customer Lifetime Value, Predictive AI, Multimodal Learning, RoBERTa, XGBoost*

JEL Classification: C45 · L83 · M31

✉ milena.nikolic@akademijanis.edu.rs
[1] Academy of Applied Technical and Preschool Studies, 18000, Niš, Serbia
[2] University of Niš, Innovation Center, Univerzitetski trg 2, 18000, Niš, Serbia
[3] Singidunum University, Danijelova 32, 11000, Belgrade, Serbia

1. INTRODUCTION

The modern hospitality industry continuously relies on data-driven decision-making to remain competitive and sustainable. The complexity of client journeys has increased with the rapid expansion of online booking platforms and review systems, which are driven by factors such as trust, loyalty, and reputation. Customer Lifetime Value (CLV) has become a central metric for strategic management, as it estimates the financial contribution of a guest throughout their relationship with a hotel. Early identification of high-value customers is crucial for optimizing pricing strategies, improving personalization, and reinforcing loyalty initiatives. Research shows that CLV frameworks provide essential guidance for modern hospitality applications, and consumer experiences directly participate in building long-term trust (Webb et al., 2022). This study extends prior research by conducting a deeper investigation into the role of CLV in the hotel industry.

CLV can be expressed either as a continuous monetary estimate or through categories that are easier to apply in practice. Table 1 illustrates three example CLV category groups, along with their associated revenue levels, booking patterns, and retention probabilities.

Table 1. Example Customer Lifetime Value categories and metrics

CLV Category	Average revenue per guest (USD)	Repeat booking rate	Retention probability	Illustrative value (USD)
Low	200	5%	0.25	< 500
Medium	600	20%	0.55	500 – 1500
High	1500	45%	0.80	> 1500

Source: Own processing

Artificial intelligence has further transformed this landscape by introducing predictive models that strengthen managerial strategies. Kabir et al. (2024) indicate that AI-driven forecasting analytics significantly improve decision quality across hotel operations, while Henriques and Pereira (2024) further confirm that embedding predictive systems into business practices enhances financial performance. These findings show that AI-driven models can forecast churn, loyalty, and spending behavior, yielding measurable benefits involving revenue growth and increased customer retention. However, much of the current research continues to focus mainly on structured booking records, often overlooking unstructured guest reviews that can carry important emotional and experiential details.

In our earlier work, we first developed custom anomaly detection strategies to strengthen the reliability of hotel reviews (Nikolić et al., 2024a) and then introduced traditional data-driven models to identify inconsistent guest feedback (Nikolić et al., 2024b). Building on this foundation, our later work progressed to deep learning techniques for predicting hotel ratings based on observed guest reviews (Nikolić et al., 2025a) and applied transformer-based methods to detect negative feedback with greater accuracy (Nikolić et al., 2025b). Collectively, these efforts achieved strong predictive accuracy (over 90%) and established a foundation for more transparent and trustworthy review analytics in hospitality research.

This paper continues along this path, extending our focus toward a hybrid AI framework that combines structured booking attributes with unstructured guest reviews. The methodology employs RoBERTa embeddings to capture contextual sentiment, while XGBoost algorithm identifies specific patterns in numerical reservation data. By integrating these complementary modalities, the model enhances the accuracy of CLV predictions and offers interpretable results through SHAP analysis, highlighting the importance of key features.

2. PREVIOUS FINDINGS

Recent studies demonstrate the central role of predictive AI in improving hotel management outcomes, especially in forecasting customer behaviors relevant to CLV. Wu and Ma (2025) reported that a hybrid neural network combined with ensemble methods achieved 81.4-82.7% accuracy on 3.2 million records obtained from 846,000 guests, while also improving customer segmentation accuracy by 47.6%. In the same study, they further linked hybrid models to measurable business impacts, including revenue increases of 16.7–23.7%, customer retention gains of 27.3%, loyalty engagement improvements exceeding 42%, and marketing ROI uplifts of 115%. Similarly, Rodrigues et al. (2025) observed recall rates above 80% for predicting cancellations, rebookings, and food package purchases. These findings suggest that intelligent forecasting models can successfully capture various phases of guest engagement across the booking journey.

Likewise, Cheng (2024) applied logistic regression, random forest, and neural networks to predict customer churn, confirming the robustness of classic and modern models put together. Furthermore, Dursun-Cengizci and Caber (2024) achieved 80% churn prediction accuracy using random forest methods on repeat customer datasets. In a large public dataset of 119,386 customers, Choi and Choi (2020) reported an overall loyalty prediction accuracy of 98.9%, with notably higher performance for first-time guests (99.43%) compared to repeat customers (81.79%). Building on these key foundations, Alsharafa et al. (2024) proved that deep neural networks can reduce forecast error to below 12% across multiple evaluation metrics (MAPE, MSE, RMSE), outperforming decision-tree and random-forest baselines.

Related studies provided by Buhalis et al. (2022) and Gatera (2024) underscore the value of integrating forecasting CLV with revenue management workflows, providing smarter pricing, reduced online travel agency commissions, and notably improved revenue per available room (RevPAR). Adhegaonkar et al. (2024) further demonstrated that automated machine learning systems can match the performance of manually optimized models on moderate-sized CLV datasets, thereby simplifying deployment for hotels lacking advanced data science teams.

Despite these advances, most existing models remain focused on structured transactional and behavioral data. While features such as recency, frequency, and monetary value remain highly predictive, they often overlook emotional and experiential cues embedded in online reviews. As Shen (2024) argues, explainable AI is progressively necessary to manage influential guest feedback and gain trust in automated decision-making. The limited integration of unstructured review records into CLV models underscores the need for hybrid frameworks that combine booking data with textual sentiment and experiential signals.

3. METHODOLOGY

The methodological framework for this study consists of a multimodal pipeline that integrates structured reservation data with unstructured guest reviews to capture both behavioral and experiential insights for predicting Customer Lifetime Value (CLV). The process unfolds in five phases: identifying data sources, preprocessing inputs, integrating RoBERTa embeddings with booking patterns extracted using XGBoost, conducting supervised model training with optimization, and evaluating performance through SHAP-based interpretability.

3.1. Data Sources

The empirical analysis relies on a multi-input dataset that combines reservation records with narrative customer feedback, providing a clear representation of behavioral transactions and subjective experiences. The structured component involved operational booking details such as stay duration, advance booking time, frequency of modifications, cancellation history, and the presence of special requests, along with guest composition (adults, children, and babies), followed by average daily rate, repeated indicators for guests, and distribution channels. The unstructured component consisted of reviews written by guests, enriched with numeric scores, reviewer metadata (including guest nationality and trip type), timestamps, and descriptions of features like service quality, cleanliness, staff interactions, and overall satisfaction.

Together, these cooperative data collections supported the integration of quantitative booking behaviors with qualitative impressions, forming a solid foundation for CLV prediction. All reviews were sourced from publicly available datasets, ensuring transparency, reliability, and the possibility of replication in future studies.

Specifically, this analysis uses the *Hotel Booking Demand Dataset*, with 119,390 reservation records across resorts and city hotels in Portugal (Antonio et al., 2019), then the *Booking.com 515K Hotel Reviews Data in Europe*, with approximately 515,000 reviews from global properties (Liu, n.d.), and the *TripAdvisor Hotel Reviews Dataset*, containing approximately 878,500 reviews with ratings and narrative feedback (Arvidsson, n.d.).

To ensure linguistic consistency in the dataset, only English-language reviews were retained. In addition, sentiment variables were derived from the obtained textual data, capturing service features like cleanliness, amenities, distance, and comfort, with each aspect assigned scores based on sentiment. Building on our prior research, inconsistent or anomalous reviews were filtered out to enhance data integrity and reduce noise in the modeling process.

Furthermore, considering that datasets did not share a common hotel identifier, the integration was performed at the feature level rather than directly mapping individual hotels. Structured reservation information from the booking dataset and sentiment-driven attributes from the reviews were combined in a unified fusion layer, enabling the model to learn from behavioral, transactional, and experiential signals without requiring a one-to-one hotel matching.

This integration strategy addresses limitations of prior CLV research, which has relied mainly on transactional records. While these capture frequency and monetary value, they overlook emotional and experiential aspects that shape loyalty. By linking booking patterns with guest narratives, this study offers a more accurate and interpretable measure of CLV that reflects both financial behavior and customer experience.

3.2 Data Preprocessing

To prepare the multimodal dataset for analysis, a series of preprocessing steps was applied. The structured reservation data required standardization and cleaning to maintain consistency across sources. Variables involving length of stay, lead time, cancellation status, and special requests were normalized into common units and formats. Missing values were imputed using median substitution for numerical features, and mode replacement for categorical features, while outliers (e.g., excessively long stays or unrealistic lead times) were discarded to reduce skewness. All

categorical features were encoded into numeric form using one-hot encoding to support their direct use in machine learning models. A summary of the datasets and variables retained after preprocessing is provided in Table 2.

Table 2. Dataset statistics before and after preprocessing

Dataset	Raw Records	Preprocessed Records	Key Features
Hotel Booking Demand	119,390	~115,000	Reservation details, guest types, dates, cancellations, modifications, special requests
Booking.com Hotel Reviews	~515,000	~475,000	Removed inconsistent reviews, retained ones with key reviewer metadata and star ratings
TripAdvisor Hotel Reviews	~878,500	~720,000	English-only reviews, aspect-based sentiments (cleanliness, amenities, distance, comfort)

Source: Own processing

Moreover, the unstructured review records required more extensive text preprocessing. Raw text was first cleaned to remove unnecessary punctuation, HTML tags, numerical strings, and special characters. Stopwords were eliminated to reduce noise, and all words were lowercased to maintain uniformity. Lemmatization was then applied to reduce words to their base forms, improving semantic consistency. Once the text was cleaned and normalized, it was tokenized and transformed into dense semantic representations using the RoBERTa model, selected for its strong contextual accuracy and robustness in domain-specific sentiment analysis. These embeddings captured underlying components in guest narratives, especially opinions about cleanliness, staff professionalism, or service quality.

After preprocessing, the structured and unstructured features were consolidated into a unified representation. The representation was used as input for forecasting CLV, with the model outputting predicted CLV scores that distinguish between low-, medium-, and high-value customers. Figure 1 illustrates the overall workflow of these steps.

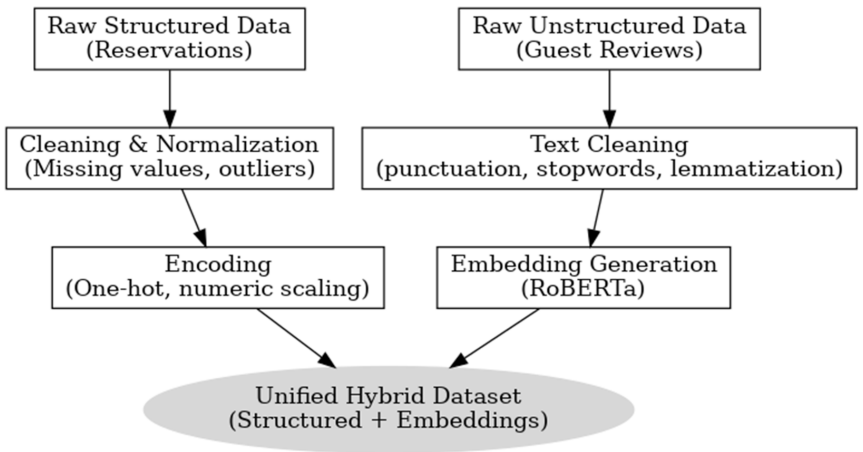

Figure 1. Preprocessing workflow for hybrid dataset construction.

Source: Own processing

3.3. Hybrid Integration

The core methodological innovation of this study lies in advancing predictive CLV modeling through the integration of heterogeneous data sources. As already highlighted in the literature review, a robust CLV prediction requires methods capable of handling complexity, ensuring interpretability, and aligning with strategic decision-making. To address these requirements, XGBoost was utilized because of its efficiency in processing high-dimensional tabular inputs, followed by abilities to model nonlinear interactions, and its resistance to overfitting, making it one of the most widely adopted algorithms for structured data (Adegoke, 2025).

In parallel, RoBERTa embeddings were introduced to represent guest experiences. As a deep transformer-based language model, RoBERTa excels at capturing context-dependent meaning by considering full sentence structure, which allows it to detect subtle shifts in sentiment (e.g., positive versus negative tones in similar phrases) and extract details from service evaluations. Unlike bag-of-words or TF–IDF representations, which consider text as unordered collections of tokens, RoBERTa retains syntactic structure and long-range dependencies between words. This capability is particularly important in hospitality reviews, where meaning frequently depends on nuanced qualifiers (e.g., *"small room but excellent service"*) and multi-aspect evaluations within the same sentence (Ibitoye et al., 2025).

In this context, RoBERTa was selected over alternatives like BERT and DistilBERT because it benefits from extended pretraining and dynamic masking, which typically demonstrate stronger performance on sentiment-based review tasks and other NLP benchmarks. DistilBERT offers greater efficiency but with lower contextual accuracy, which is critical when misclassifying high-value CLV segments carries a disproportionate managerial cost and may lead to inaccurate targeting decisions (Vijay & Premjith, 2024). Domain-specific transformer variants (including FinBERT for financial narratives and SciBERT for scientific texts) were also considered, but these models are optimized for highly specialized vocabularies and discourse patterns that differ from hospitality reviews (Tzimiris et al., 2025). Adapting them to hotel contexts without large in-domain corpora would increase the risk of overfitting and limit generalizability across hotel markets and operational environments. For these reasons, the RoBERTa model clearly provided the optimal balance between predictive accuracy and robustness required for decision-support systems in hotel management.

This fusion was implemented to establish a balanced contribution from both types of data. The resulting joint representation was then passed to the classification layer responsible for predicting CLV categories. Conceptually, the hybrid design follows a two-stream architecture in which input records are first processed independently, optimized within their respective modalities, and subsequently integrated to leverage complementary strengths.

The key rationale behind the hybrid approach is that neither behavioral variables nor textual expressions alone fully capture customer value. Guests with similar spending patterns may express very different intentions in reviews, and conversely, highly positive reviews do not always translate into high monetary value. The hybrid model therefore allows both perspectives to be represented simultaneously within the predictive process.

As noted in the data preprocessing stage, the datasets do not share common hotel identifiers, so integration is carried out at the feature level rather than by linking individual hotels or guests, which notably reduces the risk of data leakage, spurious correlations, or artificially inflated performance.

This structure encourages the hybrid model to learn generalizable relationships between booking behaviours and typical review semantics, rather than exploiting duplicated information about the same customer or property or memorizing specific patterns. This also aligns with best practices in multimodal learning, where the goal is not to perfectly reconstruct a specific dataset, but to extract cross-domain patterns that remain meaningful when applied to new hotels or market contexts.

To provide greater transparency, the hybrid model can be described in more detail as follows. In the structured-data stream, reservation attributes are provided as input to an XGBoost model that learns nonlinear relationships between behavioral indicators such as frequency of stays, length of stay, lead time, cancellations, and special requests. In the textual stream, review texts are first preprocessed and then encoded into 768-dimensional contextual embeddings using RoBERTa. These embeddings summarize semantic information relating to service quality, emotions expressed in reviews, perceived value, and guest satisfaction. In practical terms, this means that the model can account not only for *what* guests do (transactions and bookings) but also for *how they feel and what they report doing* during their stay, integrating emotional tone and perceived service quality as explicit determinants of CLV.

All preprocessing transformations and model-fitting steps are learned on the training partition only and subsequently applied to the validation and test sets, ensuring that no information from the held-out data leaks into the training process. This strict separation of training and evaluation is essential for obtaining unbiased estimates of out-of-sample performance, especially in high-capacity hybrid architectures. It also supports reproducibility, since the same preprocessing and training methods can be applied when the model is deployed on new hotels or updated datasets.

The outputs of the two streams are concatenated in a fusion layer that integrates transactional and experiential information into a unified multimodal feature vector structure. Importantly, the interaction between modalities occurs only at this fusion stage, meaning that neither stream constrains nor dominates the learning dynamics of the other. This strict separation preserves interpretability, as the effect of structured and unstructured predictors can be further inspected independently using feature-importance scores and SHAP analysis.

This design decision was intentional, as early fusion approaches can blur modality-specific effects, whereas late fusion would limit interaction across data types. The chosen intermediate-level fusion strikes a balance by enabling interaction while still allowing each modality to retain its explanatory contribution.

In practice, the fusion layer operates on a fixed-size vector obtained through concatenating the RoBERTa embedding with the latent representation learned for structured features, which is then passed to the classification head.

The fused feature vector is subsequently fed into a fully connected classification layer that outputs the probabilities of belonging to each CLV category (low, medium, high). During training, the two component models are optimized within their own domains, after which parameters of the classification head are updated based on the combined representation. This design allows the architecture to benefit from the expressiveness of deep language models and retain the stability and transparency of gradient-boosted decision trees applied to tabular data.

A detailed block-diagram architecture is provided in Figure 2, illustrating all input streams, the RoBERTa and XGBoost processing blocks, the multimodal fusion layer, and the output nodes.

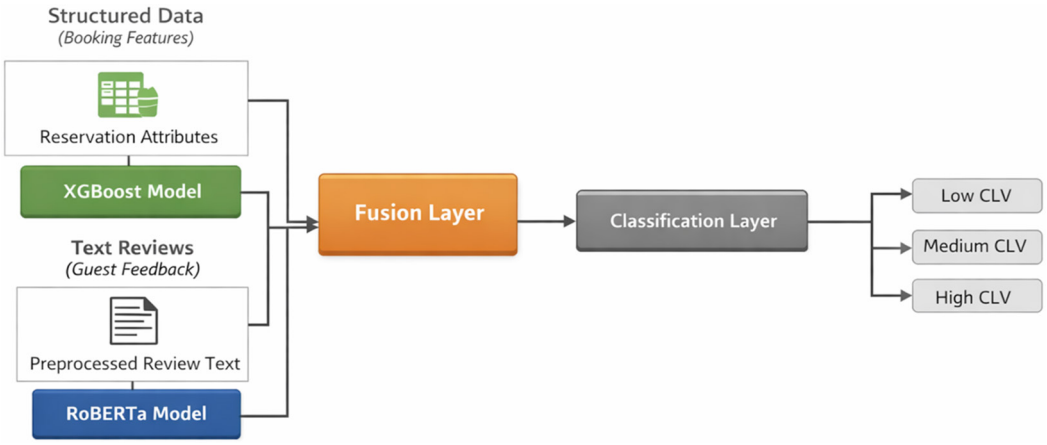

Figure 2. Hybrid CLV architecture combining XGBoost on booking features with RoBERTa review embeddings via a fusion layer and final classification layer

Source: Own processing

For clarity, the overall procedure can be summarized in pseudo-algorithmic form:

```
Input:
Structured booking variables (D_struct)
Raw review texts (D_text)
CLV labels (y)

Output:
Trained hybrid CLV model
Evaluation metrics (accuracy, precision, recall, F1-score)

Step 1: Data preparation
        Extract structured features X_struct from D_struct.
        Extract raw review texts T from D_text.
        Split (X_struct, T, y) into stratified
        training, validation, and test sets.

Step 2: Structured stream
        2.1  Preprocess X_struct (cleaning, encoding, scaling).
        2.2  Train XGBoost component on structured training data.

Step 3: Textual stream
        3.1  Clean review texts (remove noise, normalize).
        3.2  Generate RoBERTa embeddings for each review.
        3.3  Optionally compute sentiment-based features from the cleaned texts.

Step 4: Fusion layer
        4.1  For each sample, concatenate:
             structured representation from XGBoost
             textual representation from RoBERTa (and sentiment features).
        4.2  Obtain a unified multimodal feature vector Z.

Step 5: Classification head
        5.1  Train a fully connected classification layer on Z to predict CLV
             categories (low, medium, high)using the training set.
        5.2  Validate the model on the validation set and tune hyperparameters
             if required.

Step 6: Evaluation
        6.1  Apply the final hybrid model to the test set.
        6.2  Compute accuracy, precision, recall, and F1-score.
        6.3  Report performance and interpret feature contributions using SHAP
             for structured and textual predictors.
```

Taken together, the hybrid architecture operationalizes a comprehensive view of the customer, combining revenue-relevant behavioral attributes with subjective experiential assessments, which is consistent with contemporary conceptualizations of CLV in hospitality management.

As a comparative baseline, we also implemented a simpler fusion model in which sentiment scores derived from guest reviews were concatenated with structured booking features and used as input to an XGBoost classifier, allowing us to evaluate whether the proposed RoBERTa-based hybrid architecture provides value beyond conventional feature-level fusion.

3.4. Model Training and Optimization

The training process was designed to ensure that both data streams were optimized within their respective modalities before being fused into the hybrid framework. For the structured data stream, the XGBoost algorithm was trained with hyperparameter tuning, including the number of estimators, maximum tree depth, learning rate, and subsampling ratio. Additional adjustments such as column sampling and regularization terms were explored to balance bias and variance. Grid search combined with five-fold cross-validation was employed to identify parameter settings that minimized overfitting and maximized predictive accuracy.

Training of the unstructured module included fine-tuning with an Adam optimizer, an initial learning rate of 2e-5, and early stopping criteria based on validation loss. A dropout layer was applied during training to reduce overfitting and enhance the generalization capacity of the embeddings. This optimization strategy was inspired by earlier research demonstrating the effectiveness of low learning rates and dropout regularization in improving stability for text-based models (Mirabdolbaghi & Amiri, 2022).

After the two data streams were individually optimized, their outputs were concatenated in a multi-domain fusion layer. This representation was passed to a classification layer responsible for predicting Customer Lifetime Value (CLV) categories. The entire model was trained in a supervised manner using a stratified train-validation-test split technique, with 70 percent of the data used for training, 15 percent for validation, and 15 percent for testing. Stratification ensured that all CLV categories were proportionally represented in each split.

To further enhance robustness, the hybrid model was trained under multiple random seeds, and the results were averaged to reduce variance. Regularization techniques like L2 penalty were additionally incorporated into the classification layer to effectively mitigate overfitting.

3.5. Evaluation Metrics and Interpretability

The performance of the hybrid approach was evaluated with standard classification metrics: accuracy, precision, recall, and the F1-score. Accuracy provided a general measure of overall correctness, while precision and recall captured the model's ability to distinguish between high, medium, and low-value customers. The F1-score was further obtained as a balanced metric, considering the unequal distribution of CLV categories. To ensure effectiveness, results were reported on the test set and validated through repeated training with multiple random seeds.

In addition to forecasting performance, interpretability was incorporated to ensure practical relevance for hotel management. Using SHAP (SHapley Additive exPlanations), each feature was assigned a numerical contribution to CLV predictions, with the analysis underscoring structured

attributes and semantic indicators from reviews. SHAP is useful in this case since it quantifies the relative weight of features. For instance, repeat bookings might have a SHAP score of 0.35 compared to 0.12 for lead time (Meng et al., 2020). This transparency makes it clear how different factors influence the model, as further illustrated in the results section.

3.6. Reproducibility and Computing Environment

To ensure reproducibility and methodological transparency, all experiments were conducted in a specified software and hardware environment. The full pipeline, including preprocessing, embedding generation, model training, and evaluation, was implemented in Python 3.10**.** Structured-data models were developed using scikit-learn and XGBoost, while the text module employed PyTorch along with HuggingFace Transformers for RoBERTa fine-tuning.

Model development was performed on a workstation equipped with an NVIDIA RTX 3060 GPU (12 GB VRAM)**,** Intel i7 processor**,** and 32 GB RAM**.** Random seeds were fixed to 42 across libraries (NumPy, PyTorch, and XGBoost) to guarantee deterministic data splits and initialization procedures. The average RoBERTa fine-tuning run required approximately 2-2.5 hours, whereas XGBoost training completed in under three minutes per configuration.

To avoid risk of data leakage or optimistic bias, all preprocessing transformations, embedding training, and hyperparameter optimization were performed exclusively on the training partition and subsequently applied to validation and test data. Cross-dataset integration occurred strictly at the feature level, without shared identifiers across datasets, making sure that the hybrid model learned generalizable patterns rather than memorizing specific hotels or guests.

Table 3. Final hyperparameters used in model training

Component	Hyperparameter	Value
RoBERTa module	Learning rate	2e-5
	Batch size	16
	Epochs	4
	Max sequence length	256
	Optimizer	AdamW
	Dropout	0.1
	Random seed	42
XGBoost module	Number of estimators	300
	Max depth	6
	Learning rate	0.05
	Subsample	0.8
	Column subsample (colsample_bytree)	0.8
	L2 regularization (lambda)	1.0
	Random seed	42

Source: Own processing

Because CLV classes were imbalanced in the underlying data, experiments were repeated with class-weighted loss functions and stratified sampling, and results were averaged across multiple random seeds to ensure consistency of the estimates. Oversampling techniques such as SMOTE were deliberately not applied to avoid introducing synthetic samples that could distort the joint

distribution between reservation patterns and review semantics. To further control variance and avoid overfitting, hyperparameters for both components were selected through five-fold cross-validation on the training data, and the final values from experiments are presented in Table 3.

4. EXPERIMENTAL RESULTS

This section presents the empirical evaluation of the proposed hybrid framework. First, we report quantitative classification results obtained across the CLV categories and compare the hybrid model with alternative learning strategies. Performance is assessed through accuracy, precision, recall, and F1-score. Next, the managerial and practical relevance of the model is discussed through concrete application scenarios that show how predicted CLV segments can inform hotel revenue management, personalization strategies, and customer retention policies.

4.1. Predictive Performance

As a first step, a simpler fusion baseline was evaluated in which sentiment scores extracted from guest reviews were concatenated with structured booking features and classified using XGBoost. This baseline already improved performance compared with single-source models, reaching approximately 82% accuracy, 80% precision, and 79% recall**,** indicating that even basic sentiment-enhanced fusion can capture additional information beyond purely numerical reservations or text alone.

The evaluation of the proposed hybrid framework, however, demonstrates a further and more substantial improvement in predictive performance. As shown in Figure 3, the model that combines RoBERTa embeddings of guest reviews with XGBoost on structured reservation features consistently outperforms models trained on reservations or reviews only. The hybrid approach achieved approximately 89% accuracy, 87% precision, and 86% recall, whereas the model based on reviews only reached around 72% across metrics and the model based on reservations only remained close to 63%. The radar chart below highlights how the RoBERTa with XGBoost method maintains a balanced advantage in all crucial measures of estimation quality, confirming that behavioral signals from numerical reservation data and experiential cues from guest reviews are complementary in forecasting Customer Lifetime Value.

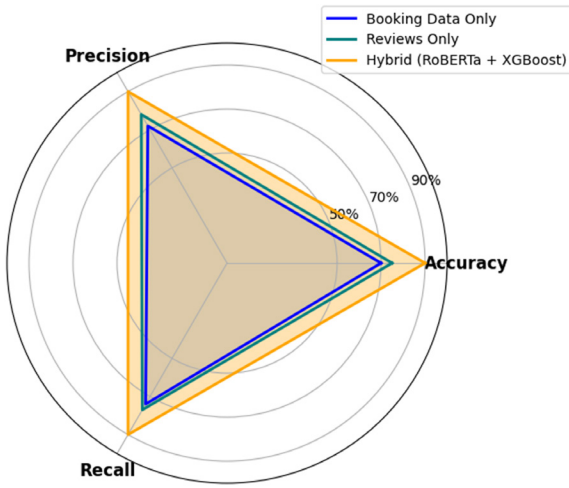

Figure 3. Performance comparison of three models across key metrics
Source: Own processing

This reduction of over 20% compared to single-source models has important implications for hotel management, since misclassifying high-value clients can result in missed opportunities for personalization, weakened loyalty programs, and inefficient resource allocation.

Beyond overall accuracy, the analysis also considered how the model manages class imbalances among low, medium, and high CLV categories. Confusion matrices and classification reports indicated that the hybrid method notably reduced the number of false negatives, particularly cases where high-value customers were misclassified as lower-value segments. Compared with the sentiment-fusion baseline, the hybrid model also achieved higher recall for the high-CLV segment, suggesting that contextual RoBERTa embeddings provide discriminative information that simple polarity scores cannot fully capture.

Interpretability was also an important component of the evaluation, as predictive models must provide transparency into the factors that influence their outcomes. SHAP (SHapley Additive exPlanations) analysis was employed to quantify feature contributions, revealing that both experiential and behavioral indicators are significant. Reviews emphasizing cleanliness, staff professionalism, and service quality were recognized as the strongest predictors of high CLV. At the same time, structured attributes like repeat bookings and special requests contributed substantially, refining the classification of guests likely to build relationships with a hotel. The results are given in Figure 4, showing the importance of structured and unstructured predictors.

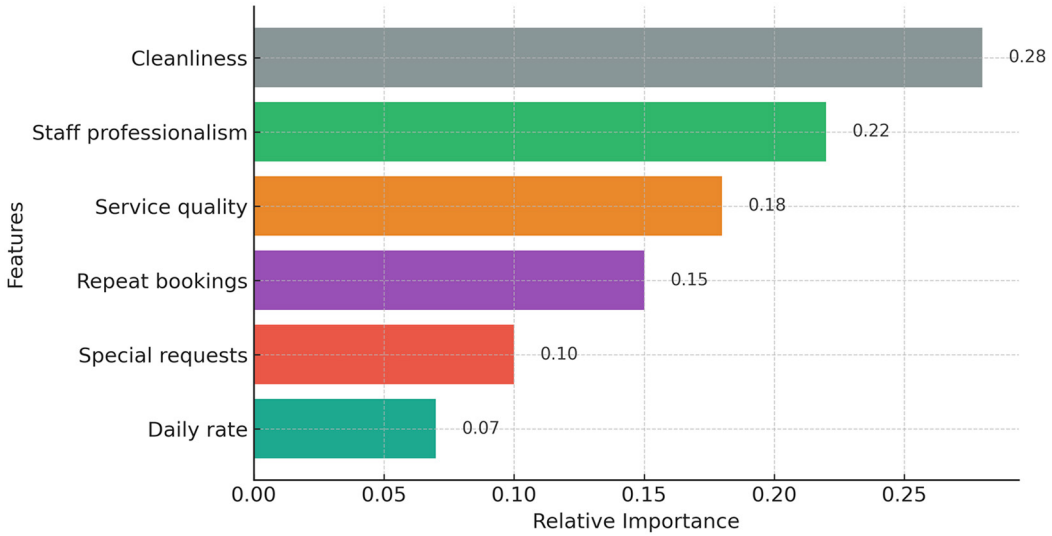

Figure 4. Relative importance of structured and unstructured features in predicting CLV
Source: Own processing

4.2. Managerial Applications

Beyond numerical improvements in classification metrics, the model has direct implications for hotel decision-making. Predicted CLV segments provide a basis for differentiated customer treatment in several operational domains. In practice, this turns CLV labels into concrete service rules. High-CLV guests identified by the hybrid model can be prioritized for proactive service recovery, complimentary upgrades, and personalized communication, increasing the likelihood of long-term retention and advocacy. Medium-CLV customers may be targeted through tailored loyalty

incentives or cross-selling strategies designed to stimulate repeat visits and gradual value escalation. Low-CLV guests can be managed using cost-efficient service configurations, helping hotels to allocate resources without compromising perceived service quality.

The hybrid model also supports data-driven segmentation strategies. By jointly analyzing transactional histories and experiential signals in reviews, hotels can distinguish, for example, between frequent but price-sensitive guests and less frequent yet high-spending customers with strong satisfaction indicators. In practice, these segment differences can be translated into clear campaign rules. These distinctions provide precise campaign design in customer relationship management systems, including personalized email marketing, pricing differentiation, and upselling of auxiliary services such as late checkout or dining packages.

Finally, the model provides early-warning capacity. Negative textual sentiment combined with previously high CLV scores may indicate guests at risk of churn, allowing hotels to intervene through tailored outreach. In this way, the framework not only classifies customers but also supports retention planning, revenue optimization, and long-term relationship management.

5. CONCLUSION

This research proposed and validated a hybrid predictive framework for forecasting Customer Lifetime Value (CLV) in the hospitality industry by aligning structured reservation data with unstructured guest reviews. By integrating behavioral and transactional indicators developed through XGBoost with contextual embeddings generated from RoBERTa, the methodology captured objective booking patterns and subjective experiential feedback. The multi-source fusion achieved higher forecasting accuracy (~89%) than models relying on either data source alone (63% and 72%), demonstrating the value of linking cooperative information streams.

The results advance both technical methodology and practical management. From a technical perspective, the results highlight how multimodal architectures can improve model robustness and interpretability, with SHAP analysis revealing the relative importance of features such as repeat bookings, length of stay, cleanliness, and service quality. From a business perspective, the proposed model enables hotels to effectively identify high-value customers and tailor engagement strategies that enhance retention and profitability.

By reframing CLV prediction as a multimodal AI task, this study shows that hotels can move from intuition-driven decisions toward systematic and evidence-based management of guest relationships, where resource allocation, personalization, and retention efforts are efficiently guided by quantified insight rather than informal judgment.

The importance of this study also lies in bridging the gap between advanced AI methods and practical decision-making in hotel management, offering a scalable and interpretable solution that can be seamlessly integrated into existing customer relationship management systems. By supporting more precise targeting and refined personalization, such strategies help hospitality providers to strengthen customer loyalty, lower churn, and maximize revenue streams. In the longer term, the adoption of hybrid AI models may fundamentally transform how hotels manage guest relationships.

Despite these contributions, the study has several limitations. The hybrid model is trained on three public datasets from specific countries and online platforms, so the learned patterns may not fully generalize to individual properties or markets with different demand structures and review

cultures. In addition, the evaluation relies on offline historical data rather than live deployment, meaning that the true business impact on revenue, loyalty, and churn still needs to be confirmed through field experiments or A/B testing.

Future research could extend this work by incorporating temporal dynamics of customer behavior, experimenting with alternate fusion strategies (e.g., attention-based mechanisms or graph-based fusion), and validating the model across diverse hospitality settings to ensure generalizability and resilience.

Acknowledgment

This research was supported by the Science Fund of the Republic of Serbia, Grant No. 7502, Intelligent Multi-Agent Control and Optimization applied to Green Buildings and Environmental Monitoring Drone Swarms – ECOSwarm.

References

Adegoke, A. (2025). Leveraging Predictive Customer Lifetime Value in Financial Consulting to Drive Growth-Focused Service Redesign Strategies. *International Journal of Research Publication and Reviews, 6*(6), 189-211. https://doi.org/10.55248/gengpi.6.0625.2220

Adhegaonkar, V. R., Thakur, A. R., Varghese, N., & Cheriyan, A. (2024, March). Decoding Customer Lifetime Value to Unlock Business Success with Predictive Machine Learning Approach. In *2024 International Conference on Trends in Quantum Computing and Emerging Business Technologies* (pp. 1-4). IEEE. https://doi.org/10.1109/tqcebt59414.2024.10545084

Alsharafa, N. S., Madhubala, P., Moorthygari, S. L., Rajapraveen, K. N., Kumar, B. R., Sengan, S., & Dadheech, P. (2024). Deep learning techniques for predicting the customer lifetime value to improve customer relationship management. *Journal of Autonomous Intelligence*, *7*(5). https://doi.org/10.32629/jai.v7i5.1622

Antonio, N., de Almeida, A., & Nunes, L. (2019). Hotel booking demand datasets. *Data in Brief, 22*, 41–49. https://doi.org/10.1016/j.dib.2018.11.126

Arvidsson, J. (n.d.). *TripAdvisor hotel reviews* [Dataset]. Kaggle. https://www.kaggle.com/datasets/joebeachcapital/hotel-reviews

Buhalis, D., O'Connor, P., & Leung, R. (2022). Smart hospitality: From smart cities and smart tourism towards agile business ecosystems in networked destinations. *International Journal of Contemporary Hospitality Management, 35*(1), 369–393. https://doi.org/10.1108/ijchm-04-2022-0497

Cheng, J. (2024). AI-Based Hotel Customer Churn Prediction Model. J*ournal of Progress in Engineering and Physical Science*, *3*(4), 15–21. https://doi.org/10.56397/jpeps.2024.12.03

Choi, Y. O., & Choi, J. (2020). The prediction of hotel customer loyalty using machine learning technique. *International Journal of Advanced Trends in Computer Science and Engineering, 9*(5), 7395–7402. https://doi.org/10.30534/ijatcse/2020/143952020

Dursun-Cengizci, A., & Caber, M. (2024). Using machine learning methods to predict future churners: An analysis of repeat hotel customers. *International Journal of Contemporary Hospitality Management, 37*(1), 36–56. https://doi.org/10.1108/ijchm-06-2023-0844

Gatera, A. (2024). Role of Artificial Intelligence in Revenue Management and Pricing Strategies in Hotels. *Journal of Modern Hospitality*, *3*(2), 14–25. https://doi.org/10.47941/jmh.1957

Henriques, H., & Pereira, L. N. (2024). Hotel demand forecasting models and methods using artificial intelligence: A systematic literature review. *Tourism & Management Studies*, *20*(3), 39-51. https://doi.org/10.18089/tms.20240304

Ibitoye, A. O. J., Kolade, O., & Onifade, O. F. W. (2025). Customer retention model using machine learning for improved user-centric quality of experience through personalised quality of service. *Journal of Business Analytics,* 1–19. https://doi.org/10.1080/2573234X.2025.2551950

Kabir, F., Khan, M. R., Mia, M. N., & Talukder, M. B. (2024). Implications of Artificial Intelligence (AI) in the Hotel Industry. In *Hotel and Travel Management in the AI Era* (pp. 357-378). IGI Global. https://doi.org/10.4018/979-8-3693-7898-4.ch017

Liu, J. (n.d.). *515K hotel reviews data in Europe* [Dataset]. Kaggle. https://www.kaggle.com/datasets/jiashenliu/515k-hotel-reviews-data-in-europe

Meng, Y., Yang, N., Qian, Z., & Zhang, G. (2020). What makes an online review more helpful: an interpretation framework using XGBoost and SHAP values. *Journal of Theoretical and Applied Electronic Commerce Research*, *16*(3), 466-490. https://doi.org/10.3390/jtaer16030029

Mirabdolbaghi, S. M. S., & Amiri, B. (2022). Model optimization analysis of customer churn prediction using machine learning algorithms with focus on feature reductions. *Discrete Dynamics in Nature and Society, 2022,* Article 5134356. https://doi.org/10.1155/2022/5134356

Nikolić, M., Stojanović, M., & Marjanović, M. (2024a). Anomaly detection in hotel reviews: Applying data science for enhanced review integrity. *2024 32nd Telecommunications Forum (TELFOR),* 1–4. https://doi.org/10.1109/TELFOR63250.2024.10819036

Nikolić, M., Stojanović, M., & Marjanović, M. (2024b). Integrating data science and predictive modeling for detecting inconsistent hotel reviews. *International Scientific Conference UNITECH 2024 – Selected Papers,* 1–16. Technical University of Gabrovo. https://doi.org/10.70456/DHXA1258

Nikolić, M., Stojanović, M., & Marjanović, M. (2025a). Integrating deep learning for automated detection of negative hotel reviews. *Facta Universitatis, Series: Automatic Control and Robotics,* 1–16. https://doi.org/10.22190/FUACR241218002N

Nikolić, M., Stojanović, M., & Marjanović, M. (2025b). The power of words: Leveraging deep learning techniques to predict hotel ratings from user reviews. *2025 24th International Symposium INFOTEH-JAHORINA (INFOTEH),* 1–6. https://doi.org/10.1109/INFOTEH64129.2025.10959201

Rodrigues, D., Jardim, B., & De Castro Neto, M. (2025). Predicting key touchpoints in hotel customer journey – a comparison of machine learning models. *Journal of Travel & Tourism Marketing, 42*(5), 609–626. https://doi.org/10.1080/10548408.2025.2456083

Shen, R. (2024). From prediction to explanation: Managing influential negative reviews through explainable AI. *arXiv.* https://arxiv.org/pdf/2412.19692v1

Tzimiris, S., Nikiforos, S., Nikiforos, M. N., Mouratidis, D., & Kermanidis, K. L. (2025). A Comparative Evaluation of Transformer-Based Language Models for Topic-Based Sentiment Analysis. *Electronics*, *14*(15), 2957. https://doi.org/10.3390/electronics14152957

Vijay, V. H., & Premjith, B. (2024). Enhancing Financial Sentiment Analysis with Pre-trained BERT-Based Models. In *Congress on Smart Computing Technologies* (pp. 79-90). Singapore: Springer Nature Singapore. https://doi.org/10.1007/978-981-96-6254-8_6

Webb, T., Cho, R., & Legg, M. P. (2022). Customer lifetime value: A data science approach for hospitality applications. *International Journal of Gaming, Hospitality and Tourism, 2*(1). https://www.ijght.org/index.php/light/article/view/46

Wu, D., & Ma, Q. (2025). AI-assisted customer behavior analysis and hotel loyalty strategy optimization. *Future Technology, 4*(3), 97–106. https://doi.org/10.55670/fpll.futech.4.3.10

ISSN: 2620-164X