

Social Risk Management of Artificial Intelligence in Higher Education: A Conceptual Framework

*Rosmi Yuhasni Mohamed Yusuf¹, Muhammad Hafiz Yaakub², Rizal Zakaria³

¹Faculty of Business and Management, Universiti Teknologi MARA, Cawangan Selangor,
Kampus Puncak Alam, Selangor, Malaysia

²Universiti Al-Quran Al-Sultan Abdullah Ahmad Shah Pahang, Malaysia

³Front Reinsurance, Malaysia

Email: *rosmi636@uitm.edu.my, mhafiz@uniqaas.edu.my, rizal@frontre.com

Corresponding Author: Dr Rosmi Yuhasni Mohamed Yusuf

Abstract: With the rapid uptake of artificial intelligence (AI) in higher education, academic risks need to be viewed as not only technical issues but as socio-institutional harms that impact equity, trust, and accountability. The methodology used for this study is a conceptual literature review and qualitative synthesis of existing academic research. This study tackles the problem existing in the literature, which is unstructured definitions, by synthesizing the literature to create a complete conceptual framework for Social Risk Management (SRM). The proposed framework places SRM in a unique role, serving as the institutional interface between the introduction of AI and social outcomes, with the role of a mediator and moderator between the raw technical inputs and the negative effects of AI on society. There are four main operational pillars of SRM identified in the synthesis. This study also adds the competency of SRM as a core competency of governance, not only for its ethical dimension but also for being a proactive competency. Moreover, it offers a clear research program for future empirical studies, detailing potential ways to measure SRM and to investigate its differences in various application areas of universities, such as admissions, financial aid, tutoring, and even assessment.

Keywords: *Social Risk Management, Artificial Intelligence, Higher Education, Institutional Governance, Conceptual Framework*

1. Introduction and Background

Artificial intelligence (AI) is becoming part of higher education to support learning analytics, tutoring, automated or assistive assessment, student support, and university governance and decision-making (Malynoshevskyi, 2023). The focus of the literature examined here is a single point: AI risks cannot be expressed as only technical problems of reliability (e.g., error rate of the model). Risks, however, are only socio-institutional harms that rely on the design, governance, and legitimization of practices enabled by AI in universities. "Risk" is then articulated as multi-dimensional, encompassing transparency, privacy, bias/discrimination, and governance, and the "social" impact of AI is then brought to the fore when systems impact participation, equity, trust, the student experience, and accountability structures.

AI initiatives in academia, as detailed by Schaeffer et al. (2024), demand governance and compliance oversight, no matter how small, in areas including transparency, user privacy, data confidentiality, data integrity, and system availability, as well as vulnerabilities like bias and discrimination. Likewise, Radzhabov et al. (2026) explicitly systemise socio-economic and social risks resulting from the use of AI in the educational process and list the social risks, which encompass digital inequality, discrimination and algorithmic bias, loss of trust and motivation. Malynoshevskyi (2023) also notes that the extensive use of AI can exacerbate existing social risks and introduce new dilemmas in the educational field, where discrimination and marginalisation are intertwined with institutional policies and implicit social norms.

This socio-technical dynamic is particularly pronounced in Malaysia, where the higher education system is experiencing a fast-paced policy-driven digital transformation (Jamaluddin et al., 2025). The transition is facilitated by strong national policies such as AI-RMAP 2021-2025 and the Digital Education Policy (DEP) that have incorporated the use of AI to develop competencies in the Fourth Industrial Revolution (IR4.0) in tertiary students (Jamaluddin et al., 2025). But it is a push from the top that has to be confronted with real-life student-level complexities. Previous empirical modelling of Malaysian undergraduates has found that performance expectancy and peer social influence (Mohamad Mozie et al., 2025) are significant factors for student adoption of generative tools such as ChatGPT. Furthermore, Abd Malik et al. (2024) found that Malaysian university

students always underestimate or overlook the critical concerns of using AI in writing, the research process, data analysis, cognitive over-dependency, and academic dishonesty.

These works stimulate the main question of this literature review: What is the relationship between social risk management and the risk connoted by AI in higher education contexts? In all studies, social risk management is defined as an institutional governance capacity that forecasts, avoids, reduces, and monitors social risks related to the adoption of AI, such as social risks relating to fairness and inclusion, human accountability, procedural legitimacy, and the maintenance of educational values. Research conceptualizing "social risk management" within the context of artificial intelligence in higher education is notably lacking.

The reviewed sources use different terms for the concept of "social risk management", but share the view that it is a system of institutional behaviours to shield stakeholders against negative societal impacts of AI. There are components that keep repeating themselves, which can be synthesized. Equity-oriented mitigation steps are needed for social risk management: discrimination, algorithmic bias, and unequal access. Digital inequality and discrimination or algorithmic bias are identified by Radzhabov et al. (2026) as a social risk. Akhtar (2024) and Ladyzhets (2024) continue this theme with an emphasis on fairness, inclusivity, and narrowing of access gaps as AI risks.

Second, social risk management requires human-centred control and accountable decision-making. Babić & Matović (2025) provide a risk-assessment logic linking probability of risk occurrence to the degree of teacher control and AI automation. Their results indicate that a combination of high control and high automation is most effective in mitigating risks, whereas low control with high automation is the riskiest scenario. This control mechanism also happens in a social way, even if the goal isn't explicitly social, as it influences the students' experience of the decisions made by AI and their ability to contest those decisions, as well as the ability of the stakeholders of the decisions to have meaningful agency.

Third, social risk management involves procedural and governance safeguards including risk-tiered authorization, impact assessment, human oversight, decision logging, audit, and contestability. Zou and Wang (2026) describe these elements in the context of an adaptive governance model of AI-powered digital governance in higher education, highlighting risk boundaries like data extraction and purpose drift, algorithmic bias and opacity, automation reliance and responsibility fragmentation, and deviation from education values. The governance orientation is also emphasized by Schaeffer et al. (2024), who consider risk-management frameworks such as NIST AI RMF and highlight stakeholder engagement, due diligence, and compliance.

Overall, the literature provides evidence that social risk management is not an isolated and marginal issue. It is an operational governance capability for organizing the use of AI towards socially acceptable outcomes and minimizing the risk of AI harms to society.

Conceptual Research Objective

The main conceptual goal of this study is to explore the connection between Social Risk Management (SRM) and the multi-dimensional risks of Artificial Intelligence (AI) adoption in Higher Education (HE). This study not only considers AI risks as technical issues of model reliability or error rates, but also as socio-institutional harms that are shaped by the design, governance, and legitimation of practices facilitated by AI in university environments. Specifically, in this study, the function of SRM as an "institutional interface" in order to capture the inputs of the AI implementation before they turn into socially consequential harms is evaluated.

2. Methodology

This study employs a Systematic Literature Review (SLR) approach to systematically gather, assess, and integrate the vast number of publications on AI deployment and governance in higher education. The review protocol is designed to be transparent, rigorous, and reproducible in literature identification and selection, following the PRISMA 2020 (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) statement and ROSES (Reporting Standards for Systematic Evidence Syntheses). To guarantee the wider and fuller coverage of higher education governance and educational technology as well as social risk management, literature retrieval was carried out via multidisciplinary and field-specific academic databases: Scopus, the

major bibliographic database, which has broad coverage of peer-reviewed journals in the field of social sciences, education, computer science, and institutional governance. Web of Science (WoS): Used to retrieve high-impact peer-reviewed research articles/letters across the Social Sciences Citation Index (SSCI) and Science Citation Index Expanded (SCI-EXPANDED). Google Scholar: A complement to the academic search engine to find the latest in-house research reports, working papers, and new press materials. Targeted search: Specialized Publisher Databases were used for targeted searches to collect specialized technical governance and education technology publications from IEEE Xplore, ScienceDirect, SpringerLink, and Wiley Online Library.

Table 1: Inclusion and Exclusion Criteria Matrix

Criterion	Inclusion Criteria	Exclusion Criteria
Timeframe	Recent literature (2023–2026) reflecting current generative AI capabilities.	Pre-2023 publications lacking relevance to current generative AI tools.
Context	Higher education institutions (undergraduate, postgraduate, administration).	K-12 primary education or corporate settings without university governance.
Thematic Focus	Social risk mitigation, institutional risk controls, equity, trust, oversight.	Purely technical software benchmarking lacking institutional social analysis.
Document Type	Peer-reviewed journal articles, indexed book chapters, conference papers.	Non-peer-reviewed blog posts, media opinion pieces, or editorials.

2. Literature Review

There are multiple ways in which AI-related risks can affect social outcomes. Several pathways are given by the reviewed sources toward the socially consequential effects of AI-related risks.

Algorithmic discrimination, inequity and participation.

The pathway that dominates is algorithmic discrimination and unequal participation. Digital inequality, falling trust and motivation, and discrimination and algorithmic bias are explicitly mentioned by Radzhabov et al. (2026) on their list of social risks. According to Malynoshevskiy (2023), there is a policy-linked explanation: discrimination and marginalisation can occur in institutional decisions based on AI, as personal values, prejudices, and implicit social norms influence social interaction logic. This means that social risk management needs to consider not just technical bias, but also institutional frameworks that license or validate the outputs from AI.

Bias is not the only equity issue. It is important to make sure that AI technologies are accessible to everyone in higher education institutions, as pointed out by Akhtar (2024), and that their use in development can help create some socio-cultural risks, like increasing inequity in access (Ladyzhets, 2024). Overall, these sources indicate that in its current form, AI is creating social risk by reconfiguring educational opportunities and exacerbating inequalities.

Trust, legitimacy and impacts on student experience

Another is the loss of trust and concerns about legitimacy. Social risks are explicitly mentioned by Radzhabov et al. (2026) and include a lack of trust and a lack of motivation. While promising, AI deployment could lead to extensive surveillance systems and management practices that can constrict the physical, emotional, and intellectual experiences of students, cautions Aper (2024). Whilst Aper's (2024) focus is on the lens of leadership decision-making, the message is that AI risk is social when it is related to student perceptions of autonomy, dignity, and perceived fairness.

From this perspective, social risk management includes managing the institutional narrative and procedural legitimacy of AI decisions. If students feel AI is interfering with their learning, taking away responsibility, or not aligned with educational values, the risks of AI are not just about the accuracy of the results but about the social impact on students, including decreased engagement, decreased motivation, and contested legitimacy.

A lack of accountability and mismatch between governance structures and accountability

A third path relates to diffusion of responsibility and misalignment of governance. Automation dependence and responsibility diffusion are the risks they outline, which can happen when AI is adopted into university governance without proper oversight and accountability structures. The risks they highlight are automation dependence and responsibility diffusion, which can arise when AI is implemented in university governance without sufficient human oversight and accountability mechanisms. They also highlight problems of misalignment with educational values and criticize the idea of AI being neutral in terms of productivity or as a standalone decision-maker. Schaeffer et al. (2024) add to that by stating the requirements for risk assessment and governance and compliance issues (e.g., transparency, privacy, data integrity, data availability). These categories are frequently referred to as “technical-governance domains” but have social implications, since the lack of transparency and accountability diffusion diminishes the power of those impacted to challenge results.

Socio-cultural transformation and human-centred educational values

Lastly, in socio-humanitarian and sociological texts, a socio-cultural path is highlighted. AI impact is placed in the context of an increasingly dangerous society by Ladyzhets (2024), who believes that social media should be coordinately managed at all levels, while at the same time emphasizing that AI gives rise to new forms of sociality in higher education. Similarly, Shahnazaryan & Shahnazaryan (2024) regard risks as increasing inequities in access and loss of human connection, while emphasizing the importance of socio-cultural and ethical considerations of AI.

All these together suggest that social risk management should see higher education as a social institution and risks as the change of learning relationships and the maintenance (or loss) of human-centred educational values.

3. Operational Pillars of Social Risk Management

Although not every source includes empirical tests of “social risk management” as a construct, there are mechanism-level mitigation approaches that, while not identical to and not all covering social risk management, can be tied to the social risk management framing.

The control-automation risk gradient and human control.

In our set, the clearest mechanism-level evidence is offered by Babić & Matović (2025). They have adopted a human-centered approach to charting risk probabilities based on teacher control and AI automation in their study. They draw the conclusion that risk reduction is best achieved with both high control and high automation, but with low control and high automation it is the riskiest combination. This finding can be socialized: decisions can be opaque or less contestable when AI automation moves forward without meaningful teacher control; teacher control is more likely to give rise to social accountability and understanding of context. This mechanism provides the foundations for a social risk management justification that institutional safeguards should not operate at its extremes: “full automation without oversight” or “control without functional integration.” Risk management, rather, relies on assigning roles in such a way that it maintains responsibility, equity, monitoring, and interpretation in the human context.

Governance frameworks: Oversight, Auditability and Contestability.

Kifarū (2026) and Zou & Wang (2026) suggest an adaptive governance framework especially developed for digital governance with AI in universities. The principles that they apply involve risk classification and authorization, impact assessment, human oversight, decision recording, audit, contestability, and institutional updating. Importantly, they do not advocate for a form of autonomous AI decision-making and do not regard AI as an automatic productivity tool. From the perspective of social risk management, these are meant to maintain the legitimacy of the procedure and diminish the risk of irreversible harm generated by AI. The same governance requirements are identified for higher education AI systems by Schaeffer et al. (2024), and the benefits of risk-management frameworks are discussed for facilitating discussion and due diligence among stakeholders. Their work is more extensive than governance in higher education institutions; however, the governance aspects are closely related to the social risk management aspects noted earlier.

Adopted responsibility strategies and ethical/inclusivity commitments

Akhtar (2024) and Tomakova & Tomakov (2026) offer integrative approaches to AI risk mitigation in education. Akhtar (2024) outlines opportunities and risks and suggests comprehensive integration strategies that have minimal impact and ethical considerations (transparency, fairness, and inclusivity). Tomakova & Tomakov (2026) list and categorize risks at the technological, ethical, psychological, cognitive, social, pedagogical, and legal/regulatory level, and state that the process of risk recognition helps to reduce negative impacts and preserve human factors through regulation, teacher retraining, and technical infrastructure. These works suggest that there is a need for social risk management to rely on both procedural governance (rules and oversight) and capability building (such as training teachers) to ensure that social safeguards can be put in place. Nagy et al. (2024) discuss the factors and obstacles towards AI being adopted in higher education systems. It brings to the fore the ethical, institutional, and geographical differences in the adoption of AI, and offers empirical grounding on the differences between institutional readiness and real risks of adoption.

Role shift in institutions and pedagogical/participatory change and reconfiguration.

Multiple sources highlight the need to reassess the roles of participants and educational approaches to address AI risks. Vidrevich et al. (2026) identify the danger of bias in evaluation, decreased critical thinking abilities, and suggest ways to mitigate the dangers through the redesign of the roles of all participants and the revision of teaching methods and pedagogical technologies. In addition, Vasoya (2023) focuses on mitigation as parental and educators' strategies and responsibilities. This transformation of roles mirrors the social aspect of risk, as it focuses on who has the authority, students' perceptions of AI interactions, and how the learning process continues to be shaped by developmental and humanistic objectives.

The social risk management-AI risk connection in higher education

Multiple risks stem from the use of AI tools in higher education: these include privacy risks, transparency risks, bias and discrimination risks, opacity risks, dependence on automation risks, and pedagogical disruption risks (Schaeffer et al., 2024; Zou & Wang, 2026; Tomakova & Tomakov, 2026; Vidrevich et al., 2026). Second, such risks associated with AI can become socially hazardous when they lead to harms that affect stakeholders' lived experience and social positions—including inequity, marginalization, loss of trust, decreased student agency, and diffusion of responsibility (Radzhabov et al., 2026; Malynoshkevskyi, 2023; Aper, 2024; Zou & Wang, 2026). Third, social risk management is built upon equity-oriented controls, human-centered oversight and governance procedures that ensure accountability, contestability and educational values, reducing AI risk (Babić & Matović, 2025; Zou & Wang, 2026; Akhtar, 2024; Malynoshkevskyi, 2023).

This synthesis indicates that social risk management can be proposed as the “institutional interface” between AI systems and social outcomes. The institutional authorization, role allocation, governance, and socio-cultural legitimacy are thus all co-productive of AI risk, as it is not simply tied to the technology itself.

4. Research Implications and Gaps

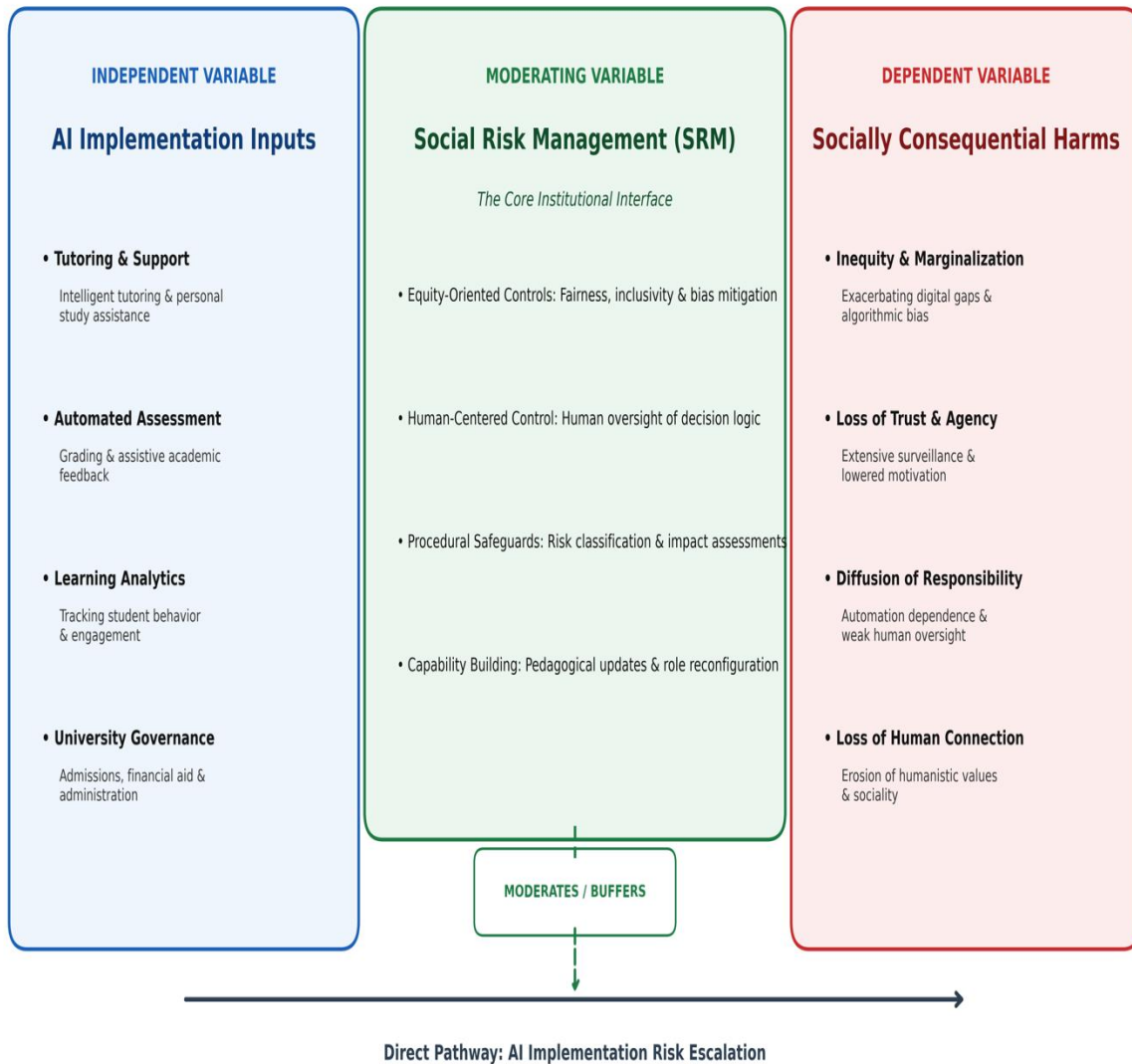
The literature reviewed offers some conceptual and governance bases that are useful, but also identifies gaps. There is a lack of uniform operational definitions of social risk management. The term “social risk management” is not defined as a measurable variable; many sources talk about social risks, governance, and ethical mitigation. This implies that more empirical research is required in order to operationalise this construct (e.g., through governance maturity, safeguards of inclusion and oversight practice, and contestability mechanisms). This is also limited by empirical causal evidence of social risk management-AI risk outcomes. While the majority give conceptual frameworks, classifications, or governance proposals, Babić & Matović (2025) provide the strongest mechanism-level quantification.

There are variable representations in higher education application domains. The correlation between social risk management and specific use cases of AI is not consistently addressed to the same extent in the risks covered, as are the other risks of admission, financial aid, tutoring, and assessment. Despite these disparities, there is a compelling theoretical argument: social risk management is an important mediator/moderator between the implementation of AI in higher education and the socially negative consequences of AI use in HE.

The conceptual framework

AI Implementation Inputs (Independent Variable): These are the major operational deployment areas, such as intelligent tutoring systems, automatic assessment and grading systems, learning analytics, and governance of the university. **Social Risk Management (Moderating Variable):** Comprises four operational pillars implemented by universities: (1) Equity-Oriented Controls, (2) Human Centred Control, (3) Procedural Safeguards, and (4) Capability Building and Role Reconfiguration. **Socially Consequential Harms (Dependent Variable):** These represent four ways in which the unchecked application of AI can disrupt the social practices of schools: algorithmic inequity and marginalization, loss of student trust and student agency, diffusion of responsibility and loss of human connection

Figure 1: The Proposed Conceptual Framework



Formal Hypothesis Specifications

H1: (Direct Baseline Effect): AI Deployment Scope has a significant positive direct relationship with Social Risk Exposure in higher education institutions.

H2: (Main Moderation Effect): Institutional Risk Controls (SRM) significantly negatively moderate the positive relationship between AI Deployment Scope and Social Risk Exposure, such that the impact of AI deployment on social risk exposure is weaker when institutional risk controls are high.

H2a: (Fairness Controls): Implementation of Fairness & Inclusion Controls significantly dampens the positive effect of AI automated evaluation on student inequity and marginalization.

H2b: (Human Oversight): Human Oversight (Human-in-the-Loop) policies significantly dampen the positive effect of AI academic assistance on student loss of trust and agency.

H2c: (Process Transparency): Process Transparency & Appeal channels significantly dampen the positive effect of administrative AI deployment on accountability confusion.

H2d: (Staff Adaptation): Staff Training & Adaptation significantly dampens the positive effect of broad AI deployment on student and faculty relational isolation.

5. Conclusion

This study reframes social risk management (SRM) as a key institutional governance function of higher education institutions (HEIs) in the fast-paced landscape of AI adoption. Instead, SRM acts as a structural moderating variable, influencing, dampening, and neutralizing the direct risk pathway between the inputs of implementing AI (analytics, automated assessment, tutoring, governance) and socially consequential harms (algorithmic inequity, loss of student trust and agency, responsibility diffusion, erosion of human connection). Universities are able to put in place an active institutional fence that prevents socio-institutional degradation by implementing a four-pillar operational framework: Equity-Oriented Controls, Human-Centered Controls, Procedural Safeguards, and Capability Building. But the need for strong SRM moderation is even greater in developing higher education ecosystems such as Malaysia, where national digital policies and the government's readiness to embrace technology converge with socio-cultural diversity. Malaysia's universities are moving into the implementation of the Malaysia National AI Roadmap, MOHE digital transformation guidelines, and institutional analytics and automation tools, with a drive toward rapid integration of generative AI. Empirical findings from recent studies in Malaysian universities reveal factors that support the use of AI tools as well as emerging issues of equity, integrity, and disciplinary variations, which align with the UTAUT model. In Malaysian higher education, the role of SRM as a moderator is different. The proposed framework puts forth SRM as a formal moderating interface that regulates the level of social risk(s) resulting from the implementation of AI in universities.

References

- Abd Malik, H. A., Azmi, N., & Tajul Arifin, S. A. (2024). *Factors influencing Malaysian university students' perceptions of artificial intelligence adoption in mathematics disciplines* (Unpublished student project). Universiti Teknologi MARA Institutional Repository. <https://ir.uitm.edu.my/id/eprint/94882>
- Akhtar, I. (2024). Artificial intelligence in higher education: Mitigating risks, embracing opportunities, and shaping the future. *Kardan Journal of Engineering and Technology*, 6(1), 1–13. <https://doi.org/10.31841/KJET.2024.35>
- Aper, J. P. (2024). Artificial intelligence and higher education: A brave new world? *Journal of Educational Thought / Revue de la Pensée Éducative*, 57(1), 3–30. <https://doi.org/10.55016/ojs/jet.v57i1.79412>
- Babić, J. Z., & Matović, S. M. (2025). Risk assessment of using artificial intelligence in education. *Sociološki Pregled [Sociological Review]*, 59(2), 610–634. <https://doi.org/10.5937/socpreg59-57266>
- Jamaluddin, F., Jamaluddin, A. H., Jamaluddin, F., & Jamaluddin, F. (2025). *Malaysia's AI-driven education landscape: Policies, applications, and comparative insights for a digital future* (arXiv Preprint arXiv:2509.21858). arXiv. <https://arxiv.org/abs/2509.21858v1>
- Kifar, F. R. (2026). Artificial intelligence tools in education systems: A structured review of applications, challenges, and implications. *Security Science Journal*, 7(1), 162–186. <https://doi.org/10.37458/ssj.7.1.8>
- Ladyzhets, N. S. (2024). Hybrid sociality: Socio-humanitarian aspects of the impact of artificial intelligence achievements on the development of higher education. *Bulletin of Udmurt University: Sociology, Political Science, International Relations*, 8(1), 42–48. <https://doi.org/10.35634/2587-9030-2024-8-1-42-48>

- Malynoshevskiy, R. (2023). Artificial intelligence (AI) and policies of social risk mitigation: The educational context. *Theoretical and Methodical Problems of Children and Youth Education*, 27(2), 25–37. <https://doi.org/10.32405/2308-3778-2023-27-2-25-37>
- Mohamad Mozie, N., Ghazali, N., & Ali Husin, L. I. (2025). Adopting artificial intelligence in higher education: Insights from the UTAUT framework on students' intentions. *Advances in Business Research International Journal*, 11(1), 71–79. <https://doi.org/10.24191/abrij.v11i1.8641>
- Nagy, A. S., Tumiwa, J. R., Arie, F. V., & Erdey, L. (2024). An exploratory study of artificial intelligence adoption in higher education. *Cogent Education*, 11(1), Article 2386892, 1–15. <https://doi.org/10.1080/2331186X.2024.2386892>
- Radzhabov, O. R., Zheligotova, R. M., & Vazkaeva, S. (2026). Socio-economic risks and limitations of implementing artificial intelligence in the educational process. *Ekonomika i Upravlenie: Problemy, Resheniya*, 5(5), 112–119. <https://doi.org/10.36871/ek.up.p.r.2026.02.05.020>
- Schaeffer, D., Coombs, L., Luckett, J., Marin, M., & Olson, P. C. (2024). Risks of AI applications used in higher education. *Electronic Journal of e-Learning*, 22(6), 60–65. <https://doi.org/10.34190/ejel.22.6.3457>
- Shahnazaryan, N. H., & Shahnazaryan, G. H. (2024). Human capital management in the context of artificial intelligence. *Region and the World*, 15(1), 78–83. <https://doi.org/10.58587/18292437-2024.1-78>
- Tomakova, R. A., & Tomakov, V. I. (2026). Artificial intelligence in higher education: Awareness of risks. *Proceedings of the Southwest State University. Series: IT Management, Computer Science, Computer Engineering. Medical Equipment Engineering*, 16(1), 8–37.
- Vasoya, N. H. (2023). The role of parents and educators in managing the risks of artificial intelligence. *Asian Journal of Education and Social Studies*, 41(4), 1–5. <https://doi.org/10.9734/ajess/2023/v41i4899>
- Vidrevich, M. B., Pervukhina, I. V., & Turovskaya, M. S. (2026). Innovations in higher education: Risks of using artificial intelligence tools. *Digital Models and Solutions*, 5(1), 44–51. <https://doi.org/10.29141/2949-477X-2026-5-1-4>
- Zou, W., & Wang, Z. (2026). Artificial intelligence-enabled digital governance in higher education: Operating mechanisms, risk boundaries, and an adaptive governance framework. *World Journal of Educational Studies*, 4(7), 62–71. <https://doi.org/10.55010/wjes-04-07-06>