

Jelica Z. Babić¹
Sofija M. Matović²

University of Belgrade, Faculty of Education,
Department of Didactics
Belgrade (Serbia)

37.015.3:159.953

37.025/.026:004.8

Original scientific paper

Submitted 05/03/2025

Revised 03/04/2025

Accepted 03/04/2025

doi: [10.5937/socpreg59-57266](https://doi.org/10.5937/socpreg59-57266)

RISKS ASSESSMENT OF USING ARTIFICIAL INTELLIGENCE IN EDUCATION³

Abstract: Recognising ethical challenges in integrating artificial intelligence in education raises many sociological questions. This study aims to identify and classify risks of using artificial intelligence in education with a quantitative assessment of the probability of risk occurrence in relation to the degree of teacher control and artificial intelligence automation according to the Human-Centred model. The theoretical analysis identified 10 risks, which were grouped into four categories. The descriptive statistical analysis based on insights from nine experts indicates 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. Future implications emphasise preventive educational measures for reliable use.

Keywords: teacher control, artificial intelligence automation, educational risks, digital society, hybrid intelligence

INTRODUCTION

Artificial Intelligence in Education (AIEd) is a multidisciplinary field that explores and applies AI technologies to enhance teaching and learning at all levels of education (Tuomi et al., 2023). In the literature, artificial intelligence in educational settings is increasingly denoted by the abbreviation AIEd, which refers to technologies designed to support teaching, learning, and decision-making processes (Hwang et al., 2020). Through the integration of various technological concepts, including machine learning, deep learning, neural networks, data analytics, and computer vision, AIEd is recognised as a powerful educational tool

¹ jelica.ristic@uf.bg.ac.rs; <https://orcid.org/0000-0003-0459-294X>

² sofija.matovic@uf.bg.ac.rs; <https://orcid.org/0009-0000-8483-8932>

³ This research was conducted by the University of Belgrade – Faculty of Education, as part of the project „Building the Critical Computer Skills for the Future-Ready Workforce” (No. 00136459) implemented by the United Nations Development Programme (UNDP) in partnership with the Ministry of Education, with the support of the Government of the Republic of Serbia.

capable of adapting to individual student needs and optimising the learning environment. Furthermore, AIED possesses the potential to facilitate the transformation of educational practices through innovative technology, facilitating personalised and collaborative learning, asynchronous discussions, and automated assessments (Nguyen et al., 2023).

Currently, AI results are limited to Artificial Narrow Intelligence, which is specialized for specific tasks such as Google Assistant, Siri, Google Translate, and other natural language processing tools. However, it is predicted that Artificial General Intelligence will possess consciousness, emotions, and genuine intelligence comparable to human intelligence (Babu & Banana, 2024). The next kind of expected AI could surpass human intelligence and lead to Artificial Super Intelligence, which will be much faster and better than humans, raising concerns about what the world will look like with such technology, along with numerous ethical dilemmas (Bostrom, 2014). Education and AI are undoubtedly interconnected. The Strategy for the Development of Artificial Intelligence in the Republic of Serbia for the Period 2025–2030 highlights the importance of education in both training AI experts and adapting to AI-driven changes in students' environments (Official Gazette of the Republic of Serbia, No. 30/18). It is clear that AI holds considerable potential to shape the future of education by opening up new possibilities for personalised learning and enhancing teaching practices, while also offering the ability to prevent discrimination, uphold fairness and justice, and avoid replacing human involvement (Schmager et al., 2023, p. 4 according to Xu et al., 2022). High-tech solutions challenge traditional pedagogical approaches by leveraging both human and artificial intelligence. AI influences the future development of education, while the modern education system contributes to its continuous progress, together promoting and enhancing technological advancement (Lufeng, 2018, p. 609).

However, AIED faces a range of challenges that go beyond the technical aspects of its implementation, including resistance within the education system. Teachers often express concerns about the mismatch between the capabilities of AI tools and the requirements of curricula, as well as ethical dilemmas and the potential erosion of their professional autonomy (Kizilcec, 2024). These challenges point to a broader social dynamic in the adoption of AI, raising not only questions about its functionality in the educational process but also about responsibility, regulation, and the ethical implications of its use in contemporary society.

The idea of AIED does not rely on unconditional trust in these systems but rather on creating an environment where teachers can gradually build confidence in the meaningful application of AI in education through experience (Kizilcec, 2024). As technology becomes increasingly integrated into education, it prompts broader sociological questions about how automation affects not only the role and autonomy of teachers, but also accountability and regulatory frameworks within the education system (Mandić et al., 2024a; Mandić et al., 2024b). The effective integration of AIED represents a complex sociotechnical challenge that necessitates a consideration of human factors to ensure their purposeful and responsible implementation (Buckingham et al., 2019). Although the number of experts capable of developing AI continues to grow, far fewer systematically analyse the implications this technology entails (Kissinger et al., 2021).

The Artificial Intelligence for K-12 initiative (AI4K12) focuses on AI's potential to improve access to education, aiming to ensure equitable access to educational resources for all students, regardless of their social or economic status, and enhance inclusivity by

reducing existing barriers (Kohnke & Zaugg, 2025; AI4K12, 2024). Ethical standards and the potential negative effects of AI in education raise important sociological questions within educational contexts.

This paper provides a detailed identification of the risks that may arise within AIEd, suggesting their classification into categories and analysing how different levels of teacher control and AI automation could have an impact on risk appearance.

Sociological dimensions of AIEd integration

The integration of AI across various societal domains raises complex questions regarding the level of system automation and its broader social implications. Examining these processes from a sociological dimension is particularly relevant, given the claims by certain scholars that contemporary society operates within a framework of surveillance capitalism (Zuboff, 2019). In this context, digital surveillance is becoming an integral part of everyday life, with a justified concern that surveillance infrastructure is being normalised and accepted without critical scrutiny (Elliot et al., 2021), just as there is a risk that AI-based automated systems may eventually be perceived as infallible, without considering the validity of their content and outcomes (Luknar, 2024a).

From a sociological standpoint, technologies are not neutral but rather shaped by socio-political and economic structures which directly influence their application and institutional regulation (Selwyn & Facer, 2014). Therefore, the integration of AI into various domains, particularly within the education system, requires carefully considered strategies grounded in ethical and socially responsible principles of digital technology usage (Elliot et al., 2021). Prospective predictions regarding the outcomes of the transition from Artificial Narrow Intelligence to Artificial General Intelligence, and subsequently to Artificial Super Intelligence, raise numerous ethical concerns, given that without clear regulatory frameworks and critical oversight, there is a risk that automated systems will not only fail to reduce existing social inequalities, but may also deepen them further.

The development of AI is an extremely dynamic process, which further complicates the creation of legal regulations capable of foreseeing the ethical framework for AI usage. This challenge becomes particularly pronounced when the level of AI automation surpasses human control, raising fundamental questions of responsibility, safety and fairness in its application (Jiang, 2023). A higher level of automation in AI systems further complicates the issue of responsibility distribution, requiring precisely defined regulatory mechanisms to ensure their lawful and socially responsible implementation while mitigating legal, ethical, and sociological risks associated with the application of AI (Elahi & Cook, 2023). While awareness of the risks and misuse of AI forms the basis for its ethical regulation, and its application contributes to human well-being, the integration of AI is justified (Luknar, 2024b). The preservation of human control in AIEd systems is a crucial prerequisite for the ethical and responsible integration of AI into the educational process, where AI can play a supportive role, but not a decisive one in pedagogical decision-making (Šuvaković, 2024).

Hybrid intelligence in the context of human control and AI automation

AI algorithms, developed through deep learning and neural networks, adapt based on human inputs, evolving to resemble human thinking. Hybrid Intelligence (HI) combines human and AI through dynamic, bidirectional interactions to achieve superior results and solve complex problems (Nguyen, 2025). The goal of HI is to leverage the strengths of both human and AI intelligence for better outcomes (Dellermann et al., 2019). Studies suggest that AI should enhance, not replace, human capabilities, particularly in teaching (Chen, 2022; Schmager et al., 2023).

In this context, HI strengthens through Human-AI Cooperation (HAC), where the teacher remains central to the learning process (Nguyen, 2025). Teachers excel in addressing innovative and non-academic issues, such as emotional understanding and creativity (Chen, 2022). AI, however, outperforms in handling structured, repetitive tasks and analysing large sets of student performance data (Zawacki-Richter et al., 2019). Although AI can offer personalised learning recommendations and feedback, it requires human oversight to ensure ethical operation (Molenaar, 2022). Human control is tied to the ability to influence actions, while automation refers to systems executing tasks independently (Alfredo et al., 2024). The balance between human control and AI automation is crucial to prevent risks from an imbalance (Domfeh et al., 2022). Molenaar suggests a framework outlining six levels of control between teachers, learners, and AI, from full teacher control to complete AI automation (Molenaar, 2022). Similarly, Nguyen identifies levels of human involvement in AI systems, ranging from “human-in-the-loop” to “human-out-of-the-loop,” with varying degrees of teacher oversight (Nguyen, 2025).

METHODOLOGY

The main concept of AIEd, in the context of hybrid intelligence and human-AI collaboration, places humans at the centre of attention. Consequently, this has directed us toward the Human-Centred AI (HCAI) approach, which can provide a solution to the ethical and purposeful integration of AI into educational environments, with the ultimate goal of improving educational outcomes, supporting students’ well-being, and preventing potential risks. Schmager et al. (Schmager et al., 2023) define HCAI as an approach that enhances human capabilities while maintaining human control over AI systems, taking into account ethical and legal aspects as well as societal well-being, and ensuring that AIEd is reliable, safe, and trustworthy, while fostering self-confidence, expertise, and creativity (Shneiderman, 2020).

This paper aims to identify and classify critical risks associated with the implementation of AIEd with quantitative assessment of the probability of risk occurrence depending on the degree of teacher control and AI automation in alignment with the Human-Centred AI (HCAI) model (Shneiderman, 2020). To improve precautions and minimise potential negative impacts, an initial theoretical analysis was carried out to identify critical risks related to AIEd. Following this, a descriptive analysis was conducted, involving the statistical processing of data obtained from sample of nine experts from diverse fields, with 2 IT specialists and 7 participants from the domain of university education, including 4

specialists in the methodology of teaching educational technologies, 1 expert in teaching methodology, 1 expert in educational sociology, 1 expert in civil security assessments.

The assessment was conducted for different modalities of human control and AI automation, represented by four quadrants (Q) of the HCAI model. Shneiderman (Shneiderman, 2022) proposed a two-dimensional framework that outlines the distinction between various degrees of human control and computer automation, emphasising that these elements are interdependent rather than mutually exclusive (Alfredo et al., 2024). More specifically, human control can be analysed in four quadrants by low or high teacher control combined with low and high automation of AIED:

Q1: LOW teacher control & LOW AIED automation: describes systems with minimal user control and negligible automation. Users can only interpret the information provided without the ability to customise or modify it, e.g., the use of learning resources in learning management systems (Alfredo et al., 2024; modified according to: Shneiderman, 2020; Shneiderman, 2022).

Q2: HIGH teacher control & LOW AIED automation: refers to systems that allow users to personalise and configure aspects of the information process, while retaining control of the learning environment with minimal or no automation of AI, e.g., personalised visualisation dashboards or educator-driven data analytics systems (Alfredo et al., 2024; modified according to: Shneiderman, 2020; Shneiderman, 2022).

Q3: LOW teacher control & HIGH AIED automation: refers to systems that rely heavily on automation with minimal user control. In these systems, algorithms manage decision-making and actions, e.g., automated grading systems that may replace or assist the teacher in specific assessment tasks (Alfredo et al., 2024; modified according to: Shneiderman, 2020; Shneiderman, 2022).

Q4: HIGH teacher control & HIGH AIED automation: describes systems that combine manual operation with automated assistance to support decision-making in learning environments. In these systems, users leverage automation to analyse learning data and make informed pedagogical decisions, e.g., intelligent teaching assistants, advanced predictive tools in analytics dashboards offering guidance and support in decision making (Alfredo et al., 2024; modified according to: Shneiderman, 2020; Shneiderman, 2022).

The research instrument was developed by intersecting a total of 10 AIED risks with different quadrants of the HCAI model in order to assess what level of teacher control and AIED automation in the educational process may influence the probability of risk occurrence. Similar risks were grouped into a single category to make the analysis more precise.

The results were collected based on the expert assessments, which were statistically analysed using mean values and standard deviations, interpreted according to the values of the measurement instrument based on a five-point Likert scale. The degree of teacher control and AI automation was compared with the assessment of the probability of risk occurrence and evaluated using the following scale:

- 1 – Strongly disagree (that risks will occur)
- 2 – Disagree to some extent (that risks will occur)
- 3 – Neither agree nor disagree (that risks will occur)
- 4 – Agree to some extent (that risks will occur)
- 5 – Strongly agree (that risks will occur)

RESULTS AND DISCUSSION

Given the complexity of AIED, a single risk can be associated with multiple aspects of AI implementation. In this context, the theoretical analysis focused on identifying the most common risks that occur during the implementation of AI in the educational context. These risks reflect general challenges that may occur during and after the implementation of AIED systems.

Classification of AIED risks

Despite its numerous advantages, the use of AI in education introduces various risks and threats that must not be ignored (Bu, 2022). To mitigate these risks, it is crucial to thoroughly identify and analyse them to prevent them and create a framework for responsible and ethical AI implementation. According to Igbokwe (Igbokwe, 2024), several potential risks (1-9) are linked to AI in education, many of which overlap with the concerns of other authors, including Yampolskiy (2022) and Zanetti et al. (2020). Additionally, some authors, such as Chan (2023) and Berendt et al. (2020) have pointed out further risks specific to the educational context. Based on the literature, the following risks emerge as key concerns in AIED implementation.

1. **Bias in AI algorithm and Decision-Making:** AI algorithms are complex, making it difficult to understand how decisions or recommendations are made. This lack of clarity can reduce trust and complicate decision-making (Igbokwe, 2024). To address this, AI systems in education require greater transparency and clearer explanations. This risk is also recognised by other authors (Bu, 2022; Yampolskiy, 2022; Zanetti et al., 2020; Čović, 2024).
2. **Data Privacy and Security** - The use of AI in education involves processing personal data, including sensitive information about students, teachers, and parents. Compliance with privacy laws, such as GDPR in Europe and FERPA in the U.S., is necessary to ensure data protection and security (Igbokwe, 2024). Ensuring data privacy and security in AI applications is a significant risk. The security of data may be compromised at various stages, increasing the risk of breaches in data security standards by AI developers (Bu, 2022; Chan, 2023).
3. **Lack of Transparency in AI Decision - (Black box)** – The lack of transparency in AI decision-making, often referred to as the “black box” effect, raises concerns about how recommendations or decisions are made (Igbokwe, 2024). Users face difficulties in understanding the processes behind AI decisions, leading to confusion and mistrust. In education, for instance, AI may suggest specific measures to address problems, but educators may not understand the factors influencing those recommendations. Other authors also emphasise the lack of transparency in AI systems (Zanetti et al., 2020).
4. **Accountability and Ethical Responsibility** - Determining responsibility becomes challenging when AI systems cause errors (Igbokwe, 2024). It is unclear whether responsibility lies with the developers of the software solutions or the educational institutions implementing them. This uncertainty creates ethical dilemmas and

- undermines trust in AI, particularly in education (Yampolskiy, 2022). AI's involvement in making critical decisions about student advancement, resource allocation, or staffing raises questions about accountability. It is necessary to clarify whether responsibility falls on the developers, educational leaders, or the institution itself (Morley et al., 2021).
5. Lack of Verifiability and Predictability - The complexity and opacity of AI decision-making processes pose significant challenges to their reliability and ethical standards, undermining trust in AI applications, especially in education. To enhance accountability and address these issues, it is crucial to increase transparency in AI operations, ensuring both effectiveness and adherence to ethical principles (Yampolskiy, 2022).
 6. Lack of Emotional Sensitivity - AI-driven systems might simulate emotions based on data, but they cannot truly understand or capture the full depth of human feelings. This lack of emotional depth could impact students' emotional development, especially if AI tools replace human interaction (Keshishi & Hack, 2023).
 7. Rigid and Limited Flexibility - AI learns independently from large amounts of data and is often unable to adapt to unpredictable situations like humans can (Holmes, 2022; Igbokwe, 2024).
 8. Risk of Dehumanisation - AI systems may shift the focus of education towards metrics and numbers, prioritising summative assessment over interpersonal relationships and personal development (Igbokwe, 2024). Education should emphasise empathy, which AI systems cannot provide, potentially weakening these essential connections within the educational system (Holmes, 2022).
 9. Loss of Autonomy for Education Managers - According to some researchers, as AI systems increasingly take over decision-making processes, educational managers may become overly dependent on these tools, potentially weakening their ability to make intuitive and context-sensitive decisions (Ahmad et al., 2023; Igbokwe, 2024).
 10. Personality Alienation - AI-powered personalised learning systems may narrow students' exposure to content only aligned with their existing interests, limiting their development and critical thinking (Bu, 2022; Chan, 2023; Ma & Jiang, 2023).
 11. Surveillance - The risk to privacy from large-scale data collection involves the potential harm of accumulating sensitive personal information (Čović, 2024). In education, technologies that gather such data may expose students, teachers, or parents to unauthorised access or misuse. The growing implementation of AI-driven systems and surveillance tools in education highlights concerns regarding personal privacy protection and compliance with relevant regulations (Berendt et al., 2020; Luknar, 2024a).
 12. Ethical Gaps' Regulation - Clear legal frameworks are essential to ensure that decisions made by AI align with ethical standards, preventing bias, mismanagement of data, and negative effects on students and teachers. This risk is particularly critical in education, where AI directly impacts individuals' lives (Berendt et al., 2020; Chan, 2023).
 13. Plagiarism - Plagiarism occurs when someone uses AI to generate content, such as text, and presents it as their work without giving credit or acknowledging that

the content was created with the help of AI. This means the person does not properly attribute the work, which violates ethical standards in education and research (Chan, 2023).

These risks in AIED can be classified into four categories based on their underlying causes: 1) sociological-pedagogical, 2) ethical, 3) technical, and 4) operational risks ([Table 1](#)). Sociological-pedagogical risks pertain to factors that may undermine the quality of the educational process, including strategic, physical, methodological, and technological elements (Kayumova & Zakirova, 2016). Ethical risks involve challenges such as systemic bias, discrimination, privacy concerns, and data misuse, which arise from the implementation of AI in educational settings (Nguyen et al., 2021). Technical risks encompass potential challenges related to the technology's performance within AIED environments (Lunesu et al., 2021), while operational risks involve disruptions caused by failures in processes, systems, or external events impacting education (Syed et al., 2021). To create the instrument evaluated by experts, out of a total of 13 identified risks, 10 were selected for assessment ([Table 2](#)).

Assessment of the probability of AIED risk occurrence

Based on the analysis of data obtained from a survey of nine experts from various disciplines, an assessment of risks within AIED was conducted, examining the probability of their occurrence depending on the balance between the degree of human control, i.e., the role of teachers, and the level of system automation. The obtained results were interpreted through four distinct quadrants, forming the basis for key research findings which are presented below.

Q1: Low teacher control & LOW AIED automation: The descriptive analysis shows that the mean values (M) for all the examined risks in the Q1 scenario were greater than or equal to 3.00, suggesting that experts largely consider there to be a certain likelihood of these risks occurring, but not to an exceptionally high extent ([Table 2](#)). The experts assess that the following risks are most likely to occur in Q1 scenarios with mean values (M) as shown in the table: 1) Lack of Verifiability and Predictability (M = 3.67), 2) Bias in algorithms and Decision-Making (M = 3.67), and 3) Lack of Emotional Sensitivity (M = 3.67). On the other hand, in conditions of low teacher and system control, the experts assess that the least likely risks to occur are: 1) Lack of Transparency in decision – (Black box) (M = 3.00), and 2) Compromised privacy through Surveillance and Monitoring (M = 3.11). The lower mean values suggest that these risks are considered less likely, probably due to the specific characteristics of the system and the possibility of better oversight or regulation of these aspects in an environment with low automation.

Q2: HIGH teacher control & LOW AIED automation: The descriptive analysis shows that the mean values (M) for all the examined risks in the Q2 scenario were significantly lower, ranging from 2.11 to 2.67 ([Table 2](#)). These values suggest that experts consider the likelihood of these risks occurring to be low in this context, which indicates that high teacher control reduces the probability of their occurrence. The highest mean values (M = 2.67) were recorded for the risks of 1) Ethical Gaps' Regulation: Plagiarism and 2) Lack of Verifiability and Predictability, indicating that experts believe these risks, although present

under conditions of Q2 scenario, still have a significant likelihood of occurring. This may be attributed to challenges related to ethical standards and the unpredictability of the system.

Q3: LOW teacher control & HIGH AIED automation: Descriptive analysis indicates that Q3 scenario is identified as the riskiest, as the mean values for each risk exceed 3.56. The highest mean values (M) within Q3 were recorded for the risks of 1) Ethical Gaps' Regulation: Plagiarism (M = 4.33) and 2) Loss of Autonomy for Education Managers (M = 3.89), reflecting significant concern among experts regarding the ethical challenges arising from high system automation (Table 2). Sociological-pedagogical risks, such as 1) Personality Alienation (M = 3.78) and 2) Risk of Dehumanisation (M = 3.78) also suggest potential negative consequences in the context of automated decision-making without adequate teacher control.

Q4: HIGH teacher control & HIGH AIED automation. The descriptive analysis shows that the mean values for nine out of ten risks in the Q4 scenario range from 1.67 to 2.33, indicating a low likelihood of their occurrence (Table 2). The only exception is the risk of Ethical Gaps' Regulation: Plagiarism, with a mean value of 2.78, suggesting concern regarding the possibility of its occurrence in this context. These results indicate that the use of AIED technologies under high teacher and system control is generally safe, except in the area of the aforementioned ethical challenge.

CONCLUSION

The separation of AIED from human cognition presents a significant challenge, requiring its conceptualisation, development, and study as integrated hybrid systems that combine human and AI (Molenaar, 2022, pp. 633–634). Achieving the right balance between human control and AI autonomy in hybrid intelligence systems remains a critical research question for experts in educational practice (Nguyen, 2025). Considering the complexity of AIED and the need to prevent negative outcomes, the identified risks are categorised into four key groups: sociological-pedagogical, ethical, technical and operational.

The research findings indicate that the probability of risk occurrence is the lowest when both teacher and system control are high, while the greatest risks arise when teacher oversight lags behind the AI system, posing significant ethical and pedagogical threats. The study results show that the Q4 scenario (HIGH teacher control & HIGH AIED automation) is the most desirable for implementing AIED technologies, as it minimises the likelihood of risk occurrence. The Q2 scenario (HIGH teacher control & LOW AIED automation) also offers relative security, but reduced automation may limit the system's efficiency in optimising the educational process, leading to increased subjectivity due to the significant influence of the teacher and the underutilisation of AI's capabilities. Although this scenario has certain advantages, its application may be limited due to the high influence of the teacher and reduced automation. The highest risks are present in the Q3 scenario (LOW teacher control & HIGH AIED automation), where high automation without adequate human supervision creates significant ethical and operational challenges, making it undesirable for implementation in AIED environments. Finally, the Q1 scenario (LOW teacher control & LOW AIED automation) also carries risks, albeit to a lesser extent than Q3. The lack

of control, even with low automation, can lead to unpredictable outcomes, making this scenario unfavourable for the implementation of AIED technologies.

From a sociological perspective, the integration of AI into the education system must be guided by ethical and socially responsible principles. In this context, to ensure the ethical application of AIED, it is crucial to establish clear legal regulations and frameworks that define ethical guidelines for the use of AI in education (Mandić, 2024; Matović & Ristić, 2024).

IMPLICATIONS

For future research and practice, integrating a structured taxonomy of hybrid intelligence into AIED could serve as a fundamental framework for advancing AI-supported education. Dellermann et al. (2019) developed a taxonomy that consolidates interdisciplinary knowledge on human-in-the-loop mechanisms in machine learning, defining key design principles for the development of such systems. Its application in the context of AIED could contribute to the creation of AI-supported educational environments that effectively balance human control and AI autonomy, thereby enhancing pedagogical outcomes, ensuring ethical and adaptive implementation, and significantly reducing the likelihood of potential risks. To prevent potential negative consequences and ensure fundamental principles of security, reliability, and trust to protect ethical practices and reduce the risk of data misuse (Alfredo et al., 2024), it is necessary to implement prevention measures related to: security protocols, data privacy protection, system reliability, collaboration with end-users, data confidentiality, and accountability in system design and implementation. The associated risks highlight the need for careful regulation as to ensure safe and ethical integration. The introduction of digital technologies in education is guided by strategic and legal frameworks that support their successful implementation (Mandić, 2023; Ristić, 2017).

Јелица З. Бабић¹
Софија М. Матовић²

Универзитет у Београду, Факултет за образовање учитеља и васпитача,
Катедра за дидактику
Београд (Србија)

ПРОЦЕНА РИЗИКА КОРИШЋЕЊА ВЕШТАЧКЕ ИНТЕЛИГЕНЦИЈЕ У ОБРАЗОВАЊУ³

(Превод *In Extenso*)

Сажетак: Препознавање етичких изазова у интеграцији вештачке интелигенције у образовање покреће бројна социолошка питања. Циљ рада односи се на идентификацију и класификацију ризика примене вештачке интелигенције у образовању, уз квантитативну процену вероватноће њихове појаве у односу на степен контроле наставника и аутоматизације вештачке интелигенције према моделу који фокус ставља на човека. Теоријском анализом идентификовано је 10 ризика груписаних у четири категорије. Дескриптивна статистичка анализа, заснована на проценама девет стручњака, указује да је комбинација високе контроле и високе аутоматизације најефикаснија у ублажавању ризика, док је ниска контрола уз високу аутоматизацију најризичнији сценарио. Будуће импликације наглашавају образовне превентивне мере поуздане употребе.

Кључне речи: контрола наставника, аутоматизација вештачке интелигенције, образовни ризици, дигитално друштво, хибридна интелигенција

УВОД

Вештачка интелигенција у образовању (енгл. AIEd) представља мултидисциплинарно поље које истражује и примењује технологије вештачке интелигенције (енгл. AI) како би унапредило наставу и учење на свим нивоима образовања (Tuomi et al., 2023). У литератури се вештачка интелигенција у образовним окружењима све више означава скраћеницом AIEd, која се односи на технологије дизајниране да подрже наставу, учење и процесе доношења одлука (Hwang et al., 2020). Интеграцијом

¹ jelica.ristic@uf.bg.ac.rs; <https://orcid.org/0000-0003-0459-294X>

² sofija.matovic@uf.bg.ac.rs; <https://orcid.org/0009-0000-8483-8932>

³ Ово истраживање спровео је Универзитет у Београду – Факултет за образовање учитеља и васпитача, у оквиру пројекта „Изградња кључних рачунарских компетенција – ка радној снази будућности” (број 00136459) који спроводи Програм Уједињених нација за развој (UNDP) у партнерству са Министарством просвете, уз подршку Владе Републике Србије.

различитих технолошких концепата, укључујући машинско и дубоко учење, неуронске мреже, анализу података и препознавање објеката, AIEd се препознаје као моћан образовни алат способан да се прилагоди индивидуалним потребама ученика и унапреди образовно окружење. Поред тога, AIEd поседује потенцијал да омогући трансформацију образовних пракси кроз иновативне технологије, олакшавајући персонализовано и сарадничко учење, асинхроне дискусије и аутоматизоване процене знања (Nguyen et al., 2023).

Тренутно су резултати вештачке интелигенције ограничени на Уску вештачку интелигенцију, која је специјализована за одређене задатке, попут Гугл асистента, Сири, Гугл преводиоца и других алата за обраду природног језика. Међутим, предвиђа се да ће Генерална вештачка интелигенција поседовати свест, емоције и праву интелигенцију упоредиву с људском интелигенцијом (Babu & Banana, 2024). Следећа врста очекиване вештачке интелигенције могла би надмашити људску интелигенцију и довести до Супер вештачке интелигенције, која ће бити много бржа и боља од људске, изазивајући забринутост о томе како ће свет изгледати са таквом технологијом, уз бројне етичке дилеме (Bostrom, 2014). Образовање и вештачка интелигенција су несумњиво повезани. Стратегија развоја вештачке интелигенције у Републици Србији за период 2025–2030. истиче значај образовања, како у погледу обуке стручњака за AI, тако и у адаптацији на промене које она доноси у окружење ученика (Službeni glasnik Republike Srbije, broj 30/18). Јасно је да AI има значајан потенцијал да обликује будућност образовања, отварајући нове могућности за персонализовано учење и унапређење наставне праксе, а истовремено нудећи алате за спречавање дискриминације, очување правичности и правде и очување људске улоге у настави (Schmager et al., 2023, стр. 4 према: Xu et al., 2022). Савремена технолошка решења се ослањају на комбинацију људске и вештачке интелигенције и као таква представљају изазов за традиционалне педагошке приступе. AI утиче на будући развој образовања, док модерни образовни системи доприносе њеном континуираном напретку, заједно подстицајући и унапређујући технолошки развој (Lufeng, 2018, стр. 609).

Међутим, AIEd се суочава са бројним изазовима који превазилазе техничке аспекте његове имплементације, укључујући и отпор унутар образовног система. Наставници често изражавају забринутост због неслагања између могућности алата заснованих на AI и захтева наставног програма, као и због етичких дилема и потенцијалног урушавања њихове професионалне аутономије (Kizilces, 2024). Ови изазови указују на ширу друштвену динамику усвајања AI, отварајући не само питања о њеној функционалности у образовном процесу већ и питања одговорности, регулације и етичких импликација њене примене у савременом друштву.

Идеја AIEd-а се не заснива на безусловном поверењу у ове системе, већ на стварању окружења у којем наставници кроз искуство постепено стичу самопоуздање за сврсисходну примену AI у образовању (Kizilces, 2024). Како технологија постаје све више интегрисана у образовање, постављају се шири социолошка питања о томе како аутоматизација утиче не само на улогу и аутономију наставника већ и на регулаторне оквири и питања одговорности унутар образовног система (Mandić et al., 2024a; Mandić et al., 2024b). Ефикасна интеграција AIEd-а представља сложен социотехнолошки изазов који захтева разматрање људских фактора како би се обезбедила његова

сврсисходна и одговорна примена (Buckingham et al., 2019). Иако број стручњака који су способни да развијају AI наставља да расте, далеко је мање оних који систематски анализирају импликације које ова технологија носи (Kissinger et al., 2021).

Иницијатива Вештачка интелигенција за K12 (AI4K12) фокусира се на потенцијал AI за побољшање приступа образовању, са циљем да обезбеди равноправан приступ образовним ресурсима за све ученике, без обзира на њихов социоекономски статус, као и да повећа инклузивност смањујући постојеће баријере (Kohnke & Zaugg, 2025; AI4K12, 2024). Етички стандарди и потенцијално негативни ефекти ВИ у образовању отварају значајна социолошка питања у оквиру образовног контекста.

Овај рад пружа детаљну идентификацију и класификацију ризика који могу настати у оквиру AIEd-a, уз анализирање различитих нивоа контроле наставника и аутоматизације AI који могу утицати на њихову појаву.

Социолошке димензије интеграције AIEd-a

Интеграција AI у различите друштвене домене отвара сложена питања у погледу нивоа аутоматизације система и њених широким социјалних импликација. Испитивање ових процеса из социолошке перспективе је посебно релевантно, имајући у виду тврдње одређених истраживача да савремено друштво функционише у оквиру надзорног капитализма (Zuboff, 2019). У том контексту, дигитални надзор постаје неодвојив део свакодневног живота, што изазива оправдану забринутост да би се инфраструктура надзора могла нормализовати и прихватити без критичке процене (Elliot et al., 2021) као што постоји ризик да се аутоматизовани системи засновани на AI временом перципирају као непогрешиви, без разматрања валидности њихових садржаја и исхода (Luknar, 2024a).

Са социолошког становишта, технологије нису неутралне, већ су обликоване социополитичким и економским структурама које директно утичу на њихову примену и институционалну регулацију (Selwyn & Facer, 2014). Стога, интеграција AI у различите области, а посебно у образовни систем, захтева пажљиво осмишљене стратегије засноване на етичким и друштвено одговорним принципима употребе дигиталних технологија (Elliot et al., 2021). Будућа предвиђања у вези са последицама преласка са Artificial Narrow Intelligence на Artificial General Intelligence, а потом и на Artificial Super Intelligence отварају бројне етичке дилеме. Уколико нема јасних регулаторних оквира и заузимања критичког става према надзору, постоји ризик да аутоматизовани системи не само што неће смањити постојеће социјалне неједнакости већ би их могли додатно продубити.

Развој AI је изузетно динамичан процес, што додатно компликује креирање правних регулатива које би могле унапред предвидети етички оквир за њену употребу. Овај изазов постаје нарочито изражен када ниво аутоматизације система AI надмашује људску контролу, чиме се покрећу фундаментална питања одговорности, безбедности и правичности њене примене (Jiang, 2023). Виши степен аутоматизације додатно усложњава питање расподеле одговорности, захтевајући прецизно дефинисане регулаторне механизме који ће осигурати закониту и друштвено одговорну примену AI, уз истовремено умањење правних, етичких и социолошких ризика који

прате њену употребу (Elahi & Cook, 2023). Док је свест о ризицима и злоупотребама AI основа за њено етичко регулисање, а њена примена доприноси добробити човека, интеграција AI је оправдана (Luknar, 2024b). Очување људске контроле у AIEd системима представља кључни предуслов за етичку и одговорну интеграцију AI у образовни процес. У том контексту, AI може имати подржавајућу улогу, али не и одлучујућу, посебно у контексту доношења педагошких одлука (Šuvaković, 2024).

Хибридна интелигенција у контексту људске контроле и аутоматизације вештачке интелигенције

Алгоритми AI, развијени кроз дубоко учење и неуронске мреже, прилагођавају се на основу људских уноса и подсећају на људско размишљање. Хибридна интелигенција (енгл. HI) комбинује људску и вештачку интелигенцију кроз динамичке, двосмерне интеракције како би се постигли изванредни резултати и решавали сложени проблеми (Nguyen, 2025). Циљ HI је да искористи предности и људске и вештачке интелигенције ради бољих исхода (Dellermann et al., 2019). Студије сугеришу да AI треба да унапреди, а не да замени људске способности, посебно у образовању (Chen, 2022; Schmager et al., 2023).

У овом контексту, HI се јача кроз сарадњу људске и вештачке интелигенције (енгл. HAI) где наставник остаје централна фигура у процесу учења (Nguyen, 2025), будући да је несумњива његова успешност у решавању иновативних и неакадемских изазова, као што су емоционално разумевање и креативност (Chen, 2022). AI, с друге стране, надмашује људе у обављању структурисаних, цикличних задатака и анализи великих скупова података о успеху ученика (Zawacki-Richter et al., 2019). Иако AI може да понуди персонализоване препоруке за учење и повратне информације, неопходан је људски надзор како би се осигурало етичко функционисање система (Molenaar, 2022). Њудска контрола односи се на способност утицања на акције, док аутоматизација подразумева системе који самостално извршавају задатке (Alfredo et al., 2024). Равнотежа између људске контроле и аутоматизације AI кључна је како би се спречили ризици услед неравнотеже (Domfeh et al., 2022). Моленар (Molenaar, 2022) предлаже оквир са шест нивоа контроле између наставника, ученика и AI, од потпуне контроле наставника до потпуне AI аутоматизације. Слично томе, Нуен (Nguyen, 2025) идентификује нивое људског укључивања у AI системе, у распону од „човека у петљи” до „човека ван петље”, са различитим степенима наставничког надзора.

МЕТОДОЛОГИЈА

Главни концепт AIEd-а, у контексту хибридне интелигенције и синергије људске и вештачке интелигенције, ставља човека у центар пажње. Сходно томе, ово нас је усмерило ка приступу вештачке интелигенције фокусиране на човека (енгл. HCAI), који може пружити решење за етичку и сврсисходну интеграцију AI у образовна окружења, са крајњим циљем унапређења образовних исхода, подршке добробити ученика и спречавања потенцијалних ризика. Шмегер и сарадници (Schmager et al., 2023) дефинишу HCAI као приступ који унапређује способности човека уз задржавање

људске контроле над системима AI, узимајући у обзир етичке и правне аспекте, као и друштвену добробит чиме се осигурава да AIEd буде поуздан, безбедан и достојан поверења, уз истовремено подстицање самопоуздања, стручности и креативности (Shneiderman, 2020).

Циљ овог рада су идентификација и класификација кључних ризика повезаних са имплементацијом AIEd-а, уз квантитативну процену вероватноће њиховог настанка у зависности од степена контроле наставника и аутоматизације AI, у складу са моделом HCAI (Shneiderman, 2020). Како би се унапредиле превентивне мере, а потенцијални негативни утицаји свели на минимум, прво је спроведена теоријска анализа ради идентификације ризика у области AIEd-а. Након тога, извршена је дескриптивна анализа која је обухватила статистичку обраду података добијених од узорка који чини девет експерата из различитих области: два стручњака за дигиталне технологије и седам из домена универзитетског образовања, укључујући четири експерта за методiku наставе образовне технологије, по једног за методiku наставе, образовну социологију и процену цивилне безбедности.

Процена је спроведена за различите модалитете људске контроле и AI аутоматизације, који су представљени кроз четири квадранта (Q) модела HCAI. Шнајдерман (Shneiderman, 2022) је предложио дводимензионални оквир који разликује различите степене људске контроле и аутоматизације система, наглашавајући да су ови елементи међусобно зависни, а не искључиви (Alfredo et al., 2024). Другим речима, људска контрола може се анализирати у четири квадранта, који комбинују низак или висок ниво наставничке контроле са ниском или високом аутоматизацијом AIEd-а:

K1: НИСКА контрола наставника и НИСКА аутоматизација AIEd – описује системе са минималном корисничком контролом и занемарљивом аутоматизацијом. Корисници могу само да тумаче доступне информације без могућности прилагођавања или модификације, на пример коришћење ресурса за учење у системима за управљање учењем (Alfredo et al., 2024; модификовано према: Shneiderman, 2020; Shneiderman, 2022).

K2: ВИСОКА контрола наставника и НИСКА аутоматизација AIEd – односи се на системе који омогућавају корисницима персонализацију и конфигурисање аспеката процеса учења, при чему наставник задржава контролу над окружењем за учење уз минималну или никакву аутоматизацију AI, на пример персонализовани визуелни панели или системи анализе података вођени од стране наставника (Alfredo et al., 2024; модификовано према: Shneiderman, 2020; Shneiderman, 2022).

K3: НИСКА контрола наставника и ВИСОКА аутоматизација AIEd – односи се на системе који се у великој мери ослањају на аутоматизацију уз минималну корисничку контролу. У овим системима, алгоритми управљају процесима одлучивања и деловања, на пример системи аутоматизованог оцењивања који могу заменити наставника или му помоћи у специфичним задацима процене (Alfredo et al., 2024; модификовано према: Shneiderman, 2020; Shneiderman, 2022).

K4: ВИСОКА контрола наставника и ВИСОКА аутоматизација AIEd – описује системе који комбинују ручно управљање са аутоматизованом подршком за доношење одлука у окружењу за учење. У овим системима, корисници користе аутоматизацију за анализу података о учењу и доношење педагошких одлука, на пример интелигентни

наставни асистенти, напредни предиктивни алати у аналитичким панелима који пружају смернице и подршку у доношењу одлука (Alfredo et al., 2024; модификовано према: Shneiderman, 2020; Shneiderman, 2022).

Истраживачки инструмент је развијен укрштањем укупно 10 AIEd ризика са различитим квадрантима модела HCAI, како би се проценило који ниво наставничке контроле и аутоматизације AIEd-а у образовном процесу може утицати на вероватноћу настанка ризика. Слични ризици су груписани у јединствене категорије ради прецизније анализе.

Резултати су прикупљени на основу експертских процена, које су статистички анализирани коришћењем средњих вредности и стандардних девијација, интерпретиране у складу са вредностима инструмента мерења на основу петостепене скале Ликертовог типа. Степен наставничке контроле и AI аутоматизације упоређен је са проценом вероватноће настанка ризика и оцењен на следећи начин:

- 1 – У потпуности се не слажем (да ће ризик настати)
- 2 – Делимично се не слажем (да ће ризик настати)
- 3 – Нити се слажем, нити се не слажем (да ће ризик настати)
- 4 – Делимично се слажем (да ће ризик настати)
- 5 – У потпуности се слажем (да ће ризик настати)

РЕЗУЛТАТИ И ДИСКУСИЈА

С обзиром на сложеност примене вештачке интелигенције у образовању (AIEd), један ризик може бити повезан са више аспеката имплементације. У том контексту, теоријска анализа је била усмерена на идентификовање најчешћих ризика који се јављају током примене вештачке интелигенције у образовном контексту. Ови ризици одражавају опште изазове који се могу јавити током и након имплементације AIEd система.

Класификација ризика узрокованих применом вештачке интелигенције у образовању

Упркос бројним предностима, примена AI у образовању доноси различите ризике и претње који се не смеју занемарити (Ви, 2022). Неопходно је детаљно идентификовати и анализирати ризике како би се спречили или ублажили, чиме би се створили услови за одговорну и етичку примену AIEd. Према Игбоквеу (Igbokwe, 2024) постоји неколико потенцијалних ризика (1–9) повезаних са AIEd, од којих се многи поклапају са резултатима студија других аутора (Yampolskiy, 2022; Zanetti et al., 2020). Поред тога, неки аутори, попут Чана (Chan, 2023), Берента и сарадника (Berendt et al., 2020) истичу специфичне ризике у образовном контексту. На основу анализе доступне литературе, следећи ризици се издвајају као кључна претња примене AIEd.

1. Пристраност у алгоритмима AI и доношењу одлука – Комплексност алгоритама отежава разумевање како се доносе одлуке или препоруке. Овај недостатак транспарентности може смањити поверење и закомпликовати процес доношења одлука (Igbokwe, 2024). У образовању је неопходна већа

- транспарентност у системима AI како би се овај проблем решио (Bu, 2022; Yampolskiy, 2022; Zanetti et al., 2020; Čović, 2024).
2. Приватност и безбедност података – Примена AI у образовању подразумева обраду личних података ученика, наставника и родитеља. Усклађеност са прописима попут GDPR у Европи и FERPA у САД је неопходна за заштиту података (Igbokwe, 2024). Присутан је ризик од цурења и злоупотребе података у различитим фазама процеса (Bu, 2022; Chan, 2023).
 3. Недостатак транспарентности у доношењу одлука (црна кутија) – Недостатак транспарентности у системима AI чини процес доношења одлука нејасним (Igbokwe, 2024). Примера ради, AI у образовању може предложити мере за решавање проблема, али наставници можда неће разумети основе тих препорука (Zanetti et al., 2020).
 4. Одговорност и етичка питања – Овај недостатак јасно дефинисане одговорности подрива поверење у AI (Yampolskiy, 2022; Morley et al., 2021). Када систем AI направи грешку, поставља се питање ко је одговоран – програмери или образовне институције (Igbokwe, 2024).
 5. Недостатак проверљивости и предвидљивости – Сложеност система AI чини њихово деловање тешко проверљивим и предвидљивим што утиче на поверење у њихову поузданост и етичке стандарде (Yampolskiy, 2022).
 6. Недостатак емоционалне осетљивости – Системи AI могу симулирати емоције али не могу истински разумети људска осећања што може негативно утицати на емоционални развој ученика када употреба алата AI мења људску интеракцију (Keshishi & Hack, 2023).
 7. Крутост и ограничена флексибилност – AI самостално учи из велике количине података и често није у могућности да се прилагоди непредвидивим ситуацијама као што то могу људи (Holmes, 2022; Igbokwe, 2024).
 8. Ризик од дехуманизације – AI може довести до тога да се образовање фокусира на бројчане метрике уместо на лични развој и емпатију (Igbokwe, 2024; Holmes, 2022).
 9. Губитак аутономије менаџера у образовању – Са све већом применом AI, менаџери у образовању могу постати претерано зависни од ових алата, што може ослабити њихову способност за доношење интуитивних и контекстуално осетљивих одлука (Ahmad et al., 2023; Igbokwe, 2024).
 10. Отуђење личности – Системи AI за персонализовано учење могу ограничити ученике на садржаје који се подударају са њиховим претходним интересовањима, ограничавајући њихов развој и критичко размишљање (Bu, 2022; Chan, 2023; Ma & Jiang, 2023).
 11. Надзор и праћење – Прикупљање великог обима података може угрозити приватност ученика, наставника и родитеља. Системи AI који користе алате за надзор могу довести до неовлашћеног приступа или злоупотребе података (Čović, 2024; Berendt et al., 2020; Luknar, 2024a).
 12. Етички јазови – Уколико не постоји јасан правни оквир, примена AIEd не би била усклађена са етичким стандардима што би довело до пристрасности и злоупотребе података (Berendt et al., 2020; Chan, 2023).

13. Плагијаризам – AI се може користити за аутоматско генерисање садржаја без адекватног навођења извора што нарушава академску етику и стандарде у образовању (Chan, 2023).

Ови ризици AIED могу се класификовати у четири категорије на основу узрока њиховог настанка: 1) социолошко-педагошки, 2) етички, 3) технички и 4) оперативни ризици (Табела 1). Социолошко-педагошки ризици односе се на факторе који могу угрозити квалитет образовног процеса, укључујући стратешке, физичке, методолошке и технолошке елементе (Kaumova & Zakirova, 2016). Етички ризици обухватају изазове као што су системска пристрасност, дискриминација, проблеми приватности и злоупотреба података који настају применом AI у образовним окружењима (Nguyen et al., 2021). Технички ризици укључују потенцијалне проблеме везане за перформансе технологије у AIED окружењима (Lunesu et al., 2021) док се оперативни ризици односе на проблеме узроковане неуспесима у процесима, системима или спољним догађајима који утичу на образовање (Syed et al., 2021). За креирање инструмента који су процењивали експерти, од укупно 13 идентификованих ризика, 10 је одабрано за евалуацију (Табела 2).

Процена вероватноће појаве ризика услед примене вештачке интелигенције у образовању

На основу анализе података добијених анкетирањем девет стручњака из различитих области, извршена је процена ризика у AIED, узимајући у обзир вероватноћу њиховог настанка у зависности од равнотеже између степена људске контроле, односно улоге наставника и нивоа аутоматизације система. Добијени резултати су интерпретирани кроз четири различита квадранта који чине основу за кључне резултате истраживања, представљене у наставку.

K1: НИСКА контрола наставника и НИСКА аутоматизација AIED – Дескриптивна анализа показује да су средње вредности (M) за све испитане ризике у сценарију K1 биле веће или једнаке 3,00 што сугерише да стручњаци углавном сматрају да постоји одређена вероватноћа појаве ових ризика, али не у изузетно високом степену (Табела 2). Као највероватније ризике у овом случају стручњаци су оценили: 1) Недостатак проверљивости и предвидивости (M = 3,67), 2) Пристрасност у алгоритму и доношењу одлука (M = 3,67) и 3) Емоционалну равнодушност (M = 3,67). С друге стране, као најмање вероватне ризике у условима ниске контроле наставника и вештачке интелигенције стручњаци су оценили: 1) Недостатак транспарентности у одлучивању (црна кутија) (M = 3,00) и 2) Угрожену приватност кроз надзор и праћење (M = 3,11). Ниже средње вредности указују на то да се ови ризици сматрају мање вероватним, вероватно због могућности боље контроле или регулације ових аспеката у окружењу са ниским степеном аутоматизације.

K2: ВИСОКА контрола наставника и НИСКА аутоматизација AIED – Дескриптивна анализа показује да су средње вредности (M) за све испитиване ризике у сценарију K2 биле знатно ниже, у распону од 2,11 до 2,67 (Табела 2). Ове вредности указују на то да стручњаци сматрају да је вероватноћа појаве ових ризика мала у овом контексту, што сугерише да висока контрола наставника смањује њихову појаву. Највише

средње вредности ($M = 2,67$) забележене су за ризике: 1) Уредба о етичким изазовима: плагијаризам и 2) Недостатак проверљивости и предвидивости, што указује да стручњаци сматрају да ови ризици, иако присутни у условима сценарија K2, и даље имају значајну вероватноћу појаве. Ово се може приписати изазовима повезаним са етичким стандардима и непредвидивошћу система.

K3: НИСКА контрола наставника и ВИСОКА аутоматизација AIEd – Дескриптивна анализа указује на то да је сценарио K3 оцењен као најризичнији, јер су средње вредности за сваки ризик веће од 3,56. Највише средње вредности (M) у оквиру K3 забележене су за ризике: 1) Уредба о етичким изазовима: плагијаризам ($M = 4,33$) и 2) Губитак аутономије за доносиоце одлука ($M = 3,89$), што одражава значајну забринутост стручњака у вези са етичким изазовима који проистичу из високе аутоматизације система (Табела 2). Социолошко-педагошки ризици, као што су: 1) Отуђење личности ($M = 3,78$) и 2) Ризик од дехуманизација ($M = 3,78$), такође указују на потенцијалне негативне последице у контексту аутоматизованог доношења одлука без адекватне контроле наставника.

K4: ВИСОКА контрола наставника и ВИСОКА аутоматизација AIEd – Дескриптивна анализа показује да средње вредности за девет од десет ризика у сценарију K4 варирају од 1,67 до 2,33, што указује на ниску вероватноћу њихове појаве (Табела 2). Једини изузетак је ризик: Уредба о етичким изазовима: плагијаризам, са средњом вредношћу од 2,78, што указује на забринутост у вези са могућношћу његове појаве у овом контексту. Резултати указују на то да је примена AIEd технологија уз високу контролу наставника и система углавном безбедна, осим у области поменутог етичког изазова.

ЗАКЉУЧАК

Одвајање AIEd од људске когниције представља значајан изазов, који захтева његову концептуализацију, развој и проучавање интегрисаних хибридних система који комбинују људску и вештачку интелигенцију (Molenaar, 2022, str. 633–634). Постизање одговарајућег баланса између људске контроле и аутономије вештачке интелигенције у хибридним интелигентним системима, остаје кључно истраживачко питање за стручњаке у области образовања (Nguyen, 2025). С обзиром на сложеност AIEd и потребу за спречавањем негативних исхода, идентификовани ризици су категорисани у четири кључне групе: социолошко-педагошке, етичке, техничке и оперативне.

Резултати истраживања указују на то да је вероватноћа појаве ризика најнижа када су и контрола наставника и контрола система високе, док се највећи ризици јављају када надзор наставника заостаје за системом AI, што представља значајне етичке и педагошке претње. Студија показује да је сценарио K4 (ВИСОКА контрола наставника и ВИСОКА аутоматизација AIEd-a) најпожељнији за имплементацију AIEd технологија јер у потпуности смањује вероватноћу појаве ризика. Сценарио K2 (ВИСОКА контрола наставника и НИСКА аутоматизација AIEd-a) такође нуди релативну сигурност, али смањена аутоматизација може ограничити ефикасност система у оптимизацији образовног процеса. У том контексту, може доћи до повећане субјективности услед значајног утицаја наставника и недовољног коришћења

могућности AI. Иако овај сценарио има одређене предности, његова примена може бити ограничена због доминације наставника и смањене аутоматизације. Највећи ризици присутни су у сценарију K3 (НИСКА контрола наставника и ВИСОКА аутоматизација AIEd-a), где висока аутоматизација без адекватног људског надзора ствара значајне етичке и оперативне изазове, чинећи га непожељним за примену у AIEd окружењима. Коначно, сценарио K1 (НИСКА контрола наставника и НИСКА аутоматизација AIEd-a) такође носи ризике, иако у мањој мери него K3. Недостатак контроле, чак и при ниској аутоматизацији, може довести до непредвидивих исхода што овај сценарио чини неповољним за имплементацију AIEd технологија.

Из социолошке перспективе, интеграција AI у образовању мора бити вођена етичким и друштвено одговорним принципима. У том контексту, како би се осигурала етичка примена AIEd-a, кључно је успоставити јасне правне регулативе и оквири који дефинишу етичке смернице за коришћење вештачке интелигенције у образовању (Mandić, 2024; Matović & Ristić, 2024).

ИМПЛИКАЦИЈЕ

За будућа истраживања и праксу, интегрисање структурисане таксономије хибридне интелигенције у AIEd може послужити као фундаментални оквир за унапређење образовања подржаног AI. Делерман и сарадници (Dellermann et al., 2019) су развили таксономију која учвршћује интердисциплинарна знања о механизмима „човека у петљи” у машинском учењу, дефинишући кључне принципе дизајна оваквих система. Њена примена у контексту AIEd могла би допринети стварању образовних окружења подржаних AI која на ефикасан начин балансирају контролу наставника и аутономију система, чиме се унапређују педагошки исходи, осигурава етичка и адаптивна примена и значајно смањује вероватноћа настанка потенцијалних ризика. Како би се спречиле могуће негативне последице и осигурали основни принципи безбедности, поузданости и поверења у циљу заштите етичких пракси и смањења ризика од злоупотребе података (Alfredo et al., 2024), неопходно је применити превентивне мере у вези са: безбедносним протоколима, заштитом приватности података, поузданошћу система, сарадњом са крајњим корисницима, поверљивошћу података и одговорношћу у дизајну и имплементацији система. Повезани ризици указују на потребу за пажљивом регулацијом како би се осигурала сигурна и етичка интеграција. Увођење дигиталних технологија у образовање треба да се одвија у оквиру стратешких и правних оквира који подржавају њихову успешну примену (Mandić, 2023; Ristić, 2017).

REFERENCES/ЛИТЕРАТУРА

- Ahmad, S. F., Han, H., Alam, M. M., Rehmat, M. K., Irshad, M., Arraño-Muñoz, M., Ariza-Montes, A. (2023). Impact of artificial intelligence on human loss in decision making, laziness and safety in education. *Humanities & Social Sciences Communication* 10 (311). <https://doi.org/10.1057/s41599-023-01787-8>

- AI4K12 (2024). Artificial intelligence (AI) for K-12 initiative (AI4K12). Available at <https://ai4k12.org/> (accessed on 27 January 2025)
- Alfredo, R., Echeverria, V., Jin, Y., Yan, L., Swiecki, Z., Gašević, D., Martinez-Maldonado, R. (2024). Human-centred learning analytics and AI in education: A systematic literature review. *Computers and Education: Artificial Intelligence* (100215). <https://doi.org/10.1016/j.caeai.2024.100215>
- Babu, M. V. S., Banana, K. (2024). A study on narrow artificial intelligence – An overview. *International Journal of Engineering Science and Advanced Technology (IJESAT)*, 24 (4), 210–219. Available at <https://www.ijesat.com/archives1.php?yr=2024&mn=4>
- Berendt, B., Littlejohn, A., Blakemore, M. (2020). AI in education: learner choice and fundamental rights. *Learning, Media and Technology*, 45 (3), 312–324. <https://doi.org/10.1080/17439884.2020.1786399>
- Bostrom, N. (2014). *Superintelligence: Paths, Dangers, Strategies*. Oxford: Oxford University Press
- Bu, Q. (2022). Ethical Risks in Integrating Artificial Intelligence into Education and Potential Countermeasures. *Science Insights*, 41(1), 561–566. <https://doi.org/10.15354/si.22.re067>
- Buckingham Shum, S., Ferguson, R., Martinez-Maldonado, R. (2019). Human-Centred Learning Analytics. *Journal of Learning Analytics*, 6 (2), 1–9. <https://doi.org/10.18608/jla.2019.62.1>
- Chan, C. K. Y. (2023). A comprehensive AI policy education framework for university teaching and learning. *International Journal of Educational Technology in Higher Education*, 20 (38). <https://doi.org/10.1186/s41239-023-00408-3>
- Chen, F. (2022). Human-AI cooperation in education: human in loop and teaching as leadership. *Journal of Educational Technology and Innovation*, 2 (1). <https://doi.org/10.61414/jeti.v2i1.34>
- Čović, A. (2024). Transformative Impact of ChatGPT Program on Education: Protection of Intellectual Property Rights and Human Rights in the Conditions of Accelerated Digitalization. In: D. Mandić, U. Šuvaković, D. V. Stanković (eds.). *Education through the COVID-19 Pandemic*, 659–672. Belgrade: Faculty of education. https://doi.org/10.18485/uf_edu_covid19.2024.2.ch21
- Dellermann, D., Calma, A., Lipusch, N., Weber, T., Weigel, S., Ebel, P. (2019). The future of human-ai collaboration: a taxonomy of design knowledge for hybrid intelligence systems. In: T. Buid (ed.). *Hawaii International Conference on System Sciences (HICSS)*. Maui, Hawaii, USA: AIS <https://doi.org/10.24251/HICSS.2019.034>
- Domfeh, E. A., Weyori, B. A., Appiahene, P., Mensah, J., Awarayi, N. S. (2022). Human-Centered Artificial Intelligence, a review. *TechRxiv*. <https://doi.org/10.36227/techrxiv.19174772.v1>
- Elahi, A., Cook, T. S. (2023). Artificial Intelligence Governance and Strategic Planning: How We Do It. *Journal of the American College of Radiology: JACR*, 20 (9), 825–827. <https://doi.org/10.1016/j.jacr.2023.06.017>
- Elliott, K., Price, R., Shaw, P., Spiliotopoulos, T., Ng, M., Coopamootoo, K., Van Moorsel, A. (2021). Towards an Equitable Digital Society: Artificial Intelligence (AI) and Corporate Digital Responsibility (CDR). *Society*, (58), 179–188. <https://doi.org/10.1007/s12115-021-00594-8>

- Holmes, W., Luckin, R., Griffiths, M. (2022). *Artificial intelligence in education: Promise and implications for teaching and learning*. Boston, USA: CCR. Available at <https://curriculumredesign.org/our-work/artificial-intelligence-in-education/>
- Hwang, G. J., Xie, H., Wah, B. W., Gašević, D. (2020). Vision, challenges, roles and research issues of Artificial Intelligence in Education. *Computers and Education: Artificial Intelligence*, (1, 100001). <https://doi.org/10.1016/j.caeai.2020.100001>
- Igbokwe, I. C. (2024). Artificial Intelligence in Educational Leadership: Risks and Responsibilities. *European Journal of Arts, Humanities and Social Sciences*, 1 (6), 3–10. [https://doi.org/10.59324/ejahss.2024.1\(6\).01](https://doi.org/10.59324/ejahss.2024.1(6).01)
- Jiang, L. (2023). Legal Risks and Responses to Artificial Intelligence Technology from the Perspective of Sociology of Technology. *Academic Journal of Management and Social Sciences*, 5 (3), 183–186. <https://doi.org/10.54097/oosnmpmt>
- Kayumova, L. R., Zakirova, V. G. (2016). Educational Environment Risks: Problems of Identification and Classification. *International Journal of Environmental and Science Education*, 11 (6), 1013–1019. <https://doi.org/10.12973/ijese.2016.370a>
- Keshishi, N., Hack, S. (2023). Emotional intelligence in the digital age: Harnessing AI for students' inner development. *Journal of Perspectives in Applied Academic Practice*, 11 (3). <https://doi.org/10.56433/jpaap.v11i3.579>
- Kissinger, H., Schmidt, E., Huttenlocher, D. (2021). *The Age of A.I. And Our Human Future*. New York: Little Brown and Company
- Kizilcec, R. (2024). To Advance AI Use in Education, Focus on Understanding Educators. *International Journal of Artificial Intelligence in Education*, 34 (1), 12–19. <https://doi.org/10.1007/s40593-023-00351-4>
- Kohnke, S., Zaugg, T. (2025). Artificial Intelligence: An Untapped Opportunity for Equity and Access in STEM Education. *Education Sciences*, 15 (1). <https://doi.org/10.3390/educsci15010068>
- Lufeng, H. (2018). Analysis of New Advances in the Application of Artificial Intelligence to Education. In: M. Pavlova, R. He (eds.). *3rd International Conference on Education, E-learning and Management Technology (EEMT 2018)*, 608–611. Bangkok, Thailand: Atlantis Press. <https://doi.org/10.2991/iceemt-18.2018.118>
- Luknar, I. (2024a). Artificial intelligence as a challenge. *Napredak – časopis za političku teoriju i praksu*, 5 (2), 17–24. <https://doi.org/10.5937/napredak5-52577>
- Luknar, I. (2024b). *Artificial intelligence, possibility or challenge*. Beograd: Medija centar Odbrana [In Serbian]
- Lunesu, M. I., Tonelli, R., Marchesi, L., Marchesi, M. (2021). Assessing the Risk of Software Development in Agile Methodologies Using Simulation. *IEEE Access*, (9) 134240–134258. <https://doi.org/10.1109/ACCESS.2021.3115941>
- Ma, X., Jiang, C. (2023). On the Ethical Risks of Artificial Intelligence Applications in Education and Its Avoidance Strategies. *Journal of Education, Humanities and Social Sciences* (14), 354–359. <https://doi.org/10.54097/ehss.v14i.8868>
- Mandić, D. (2023). Report on Smart Education in the Republic of Serbia. In: R. Zhuang, L. Dejian, D. Sampson, D. Mandić, S. Zou, Y. Huang, R. Huang (eds.). *Smart Education in China and Central & Eastern European Countries*. Singapore: Springer. https://doi.org/10.1007/978-981-19-7319-2_11

- Mandić, D. (2024). A New Paradigm of Education and the Potentials of Artificial Intelligence. *Napredak – časopis za političku teoriju i praksu*, 5 (2), 83–96. <https://doi.org/10.5937/napredak5-51939> [In Serbian]
- Mandić, D., Mišćević, G., Bujišić, Lj. (2024a). Evaluating the quality of responses generated by ChatGPT. *Metodička teorija i praksa*, 27 (1), 5–19. <https://doi.org/10.5937/metpra27-51446>
- Mandić, D., Mišćević, G., Babić, J., Matović, S. (2024b). Educational robots in teachers' education. *Istraživanja u pedagogiji*, 14 (2), 361–376. <https://doi.org/10.5937/IstrPed2402361M>
- Matović, S., Ristić, M. (2024). Educational robots in the function of developing algorithmic thinking of preschool children. *Metodička teorija i praksa*, 27 (2), 21–35. <https://doi.org/10.5937/metpra27-52258> [In Serbian]
- Molenaar, I. (2022). Towards hybrid human-AI learning technologies. *European Journal of Education*, 57 (4), 632–645. <https://doi.org/10.1111/ejed.12527>
- Morley, J., Floridi, L., Kinsey, L., Elhalal, A. (2021). From what to how: An initial review of publicly available AI ethics tools, methods and research to translate principles into practices. *Ethics and Information Technology*, 23 (2), 299–316. <https://doi.org/10.1007/s10676-020-09545-5>
- Nguyen, A. (2025). Human-AI Shared Regulation for Hybrid Intelligence in Learning and Teaching: Conceptual Domain, Ontological Foundations, Propositions, and Implications for Research. In: T. X. Bui (ed.). *58th Hawaii International Conference on System Sciences 2025.*, 34–43. Hawaii: University of Hawaii at Manoa. <https://doi.org/10.24251/HICSS.2025.006>
- Nguyen, A., Ngo, H. N., Hong, Y. et al. (2023). Ethical principles for artificial intelligence in education. *Education and Information Technologies* 28, 4221–4241. <https://doi.org/10.1007/s10639-022-11316-w>
- Ristić, M. (2017). E-Maturity in Schools. *Croatian Journal of Education*, 19 (Sp.Ed.3), 317–334. <https://doi.org/10.15516/cje.v19i0.3100> [In Serbian]
- Schmager, S., Pappas, I., Vassilakopoulou, P. (2023). Defining human-centered AI: a comprehensive review of HCAI literature. In: A. Arenas, M. I. Merhi, S. Skandrani, J. Wareham (eds.). *15th Mediterranean Conference on Information Systems and the 6th Middle East & North Africa Conference on digital Information Systems*. Madrid: IE University. Available at <https://easychair.org/publications/preprint/cGmc/open>
- Selwyn, N., Facer, K. (2014). The sociology of education and digital technology: Past, present and future. *Oxford Review of Education*, 40 (4), 482–496. <https://doi.org/10.1080/03054985.2014.933005>
- Shneiderman, B. (2020). Human-centered artificial intelligence: Reliable, safe & trustworthy. *International Journal of Human-Computer Interaction*, 36 (6), 495–504. <https://doi.org/10.48550/arXiv.2002.04087>
- Shneiderman, B. (2022). *Human-centered AI*. Oxford, New York: Oxford University Press
- Strategy for the Development of Artificial Intelligence in the Republic of Serbia for the Period 2025–2030: Official Gazette of the RS No. 30/18. (accessed on 27 January 2025), Available at https://www.media.srbija.gov.rs/medsrp/dokumenti/strategy_artificial_intelligence.pdf

- Šuvaković, U. (2024). Creator of the present in an attempt to understand the future or what can be counted as, but not reduced to, Kissinger's legacy. *Napredak – časopis za političku teoriju i praksu*, 5 (2), 129–150. <https://doi.org/10.5937/napredak5-52859>
- Tuomi, I., Cachia, R., Villar-Onrubia, D. (2023). *On the futures of technology in education: emerging trends and policy implications*. Joint Research Center, European Union. <https://doi.org/10.2760/079734>
- Xu, W., Dainoff, M. J., Ge, L., Gao, Z. (2022). Transitioning to human interaction with AI systems: New challenges and opportunities for HCI professionals to enable human-centered AI. *International Journal of Human-Computer Interaction*, 39 (3), 494–518. <https://doi.org/10.48550/arXiv.2105.05424>
- Yampolskiy, R. V. (2022). On the Controllability of Artificial Intelligence: An Analysis of Limitations. *Journal of Cyber Security and Mobility*, 11 (03), 321–404. <https://doi.org/10.13052/jcsm2245-1439.1132>
- Zanetti, M., Rendina, S., Picci, L., Peluso Cassese, F. (2020). Potential risks of Artificial Intelligence in education. *Form@re – Open Journal Per La Formazione in Rete*, 20 (1), 368–378. <https://doi.org/10.13128/form-8113>
- Zawacki-Richter, O., Marín, V. I., Bond, M., Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education—where are the educators?. *International Journal of Educational Technology in Higher Education*, 16 (1), 1–27. <https://doi.org/10.1186/s41239-019-0171-0>
- Zuboff, S. (2019). *The Age of Surveillance Capitalism: The Fight for the Future of the New Frontier of Power*. London: Profile Books Ltd

APPENDIX/ПРИЛОГ

Table 1: *Classification of identified risks of artificial intelligence in education*
/ Табела 1: Класификација идентификованих ризика вештачке
интелигенције у образовању

CATEGORY/КАТЕГОРИЈА	RISKS/РИЗИЦИ
Sociological-pedagogical risks / Социолошко-педагошки ризичи	• Lack of Emotional Sensitivity / Емоционална равнодушност • Risk of Dehumanisation / Ризик од дехуманизације • Personality Alienation / Отуђење личности
Ethical risks / Етички ризичи	• Bias in algorithm and Decision-Making / Пристрасност у алгоритму и доношењу одлука • Compromised privacy through Surveillance and Monitoring / Угрожена приватност кроз надзор и праћење • Ethical Gaps' Regulation: Plagiarism / Уредба о етичким изазовима: плагијаризам

← НАЗАД

← BACK

Technical risks / Технички ризици	• Lack of Transparency in decision – (Black box) / Недостатак транспарентности у одлучивању – (црна кутија) • Lack of Verifiability and Predictability / Недостатак проверљивости и предвидљивости • Rigid and Limited Flexibility / Ригидност и ограничена флексибилност
Operational risks / Операционални ризици	• Loss of Autonomy for Education Managers / Губитак аутономије за доносиоце одлука

Table 2: *Results of descriptive statistics on the probability of risk occurrence when using artificial intelligence in education, depending on the degree of teacher control and system automation* / Табела 2: *Резултати дескриптивне статистице о вероватноћи појаве ризика приликом коришћења вештачке интелигенције у образовању у зависности од степена контроле наставника и аутоматизације система*

	Q1/K1		Q2/K2		Q3/K3		Q4/K4		N
	M	Std. Dev.	M	Std. Dev.	M	Std. Dev.	M	Std. Dev.	
Lack of Emotional Sensitivity / Емоционална равнодушност	3,67	1,581	2,11	,928	3,67	1,225	1,67	,707	9
Risk of Dehumanisation / Ризик од дехуманизације	3,22	1,302	2,11	1,167	3,78	1,202	2,22	1,093	9
Personality Alienation / Отуђење личности	3,33	1,323	2,33	,707	3,78	1,093	2,00	,866	9
Bias in algorithm and Decision-Making / Пристрасност у алгоритму и доношењу одлука	3,67	1,658	2,44	1,236	3,56	1,014	1,67	,500	9
Compromised privacy through Surveillance and Monitoring / Угрожена приватност кроз надзор и праћење	3,11	1,364	2,56	,882	3,67	1,500	2,33	1,225	9
Ethical Gaps' Regulation: Plagiarism / Уредба о етичким изазовима: плагијаризам	3,56	1,424	2,67	1,118	4,33	1,118	2,78	,972	9
Lack of Transparency in decision – (Black box) / Недостатак транспарентности у одлучивању – (црна кутија)	3,00	1,323	2,33	1,000	3,67	1,414	2,11	1,054	9
Lack of Verifiability and Predictability / Недостатак проверљивости и предвидљивости	3,67	1,225	2,67	1,000	3,78	,833	2,22	1,093	9
Rigid and Limited Flexibility / Ригидност и ограничена флексибилност	3,44	1,509	2,11	,782	3,78	,972	2,11	,928	9
Loss of Autonomy for Education Managers / Губитак аутономије за доносиоце одлука	3,56	1,236	2,56	1,014	3,89	1,167	2,11	1,054	9

← НАЗАД

← ВАСК