Frowin Rabanus Kifaru
Faculty of Business and Information Sciences, Moshi Cooperative University, Tanzania
E-mail: frowin2005@email.com

DOI: https://doi.org/10.37458/ssj.7.1.8
Original Research Article
Received: February 16, 2026
Accepted: March 23, 2026

ARTIFICIAL INTELLIGENCE TOOLS IN EDUCATION SYSTEMS: A STRUCTURED REVIEW OF APPLICATIONS, CHALLENGES, AND IMPLICATIONS

Abstract: *Artificial intelligence (AI) is increasingly transforming higher education through applications in teaching, assessment, research, and institutional management. However, existing studies remain fragmented and often overlook governance, security risks, and ethical implications. This study presents a systematic review of AI tools in higher education from a security science perspective. Using PRISMA 2020 guidelines, peer-reviewed studies published between 2020 and 2025 were analyzed through thematic synthesis. The findings identify four major categories of AI tools: generative AI, learning analytics systems, intelligent tutoring systems, and administrative decision-support tools. While these technologies enhance efficiency and personalization, they introduce risks related to academic integrity, data privacy, algorithmic opacity, and system dependency. To address this gap, the study proposes a weighted Confidentiality–Integrity–Availability (CIA) model for quantifying AI-related risks, alongside a governance framework for institutional risk management. The results emphasize that effective AI adoption requires robust governance, ethical safeguards, and human-centered oversight. The study contributes a structured and measurable approach to evaluating AI risks in higher education systems.*

Keywords: *Artificial intelligence, Machine Learning, Higher education systems, Generative AI, Academic integrity, Security science, Data protection, Systematic review, Cybersecurity*

1. Introduction

1.1 Background

Artificial intelligence has brought significant advantages to education and has gained attention in recent decades. It has advanced significantly and been enhanced in various areas of education, including improved teaching and learning processes, school management, and improvement of overall educational accessibility (Chiu, Xia, Zhou, Chai, & Cheng, 2023). Advances in machine learning have enabled sophisticated technologies for generating digital content. Evolution of AI, which uses deep learning and machine learning, has brought the education sector into a new era by producing new digital content such as video, images, text, and audio by analyzing training data and identifying patterns and distributions (Akmeşe, Kör, & Erbay, 2021). The introduction of AI has helped universities improve teaching and learning, and support lesson planning (Tan, Cheng, & Ling, 2025). Advancements in machine learning, natural language processing, and generative AI models have widely enhanced the range of AI applications not only in higher education but also in some of the social and economic sectors (Bahroun, Anane, Ahmed, & Zacca, 2023). Many educational institutions have largely adopted artificial intelligence tools to accelerate their digitization. Also, the expansion of online learning as a result of generative systems capable of producing human-like text has been raised as well (Ward, Bhati, Neha, & Guercio, 2024).

Existing literature suggests that academic institutions are becoming overly reliant on artificial intelligence tools, which have a significant capacity to improve all education systems (Cordero, Torres-Zambrano, & Cordero-Castillo, 2024), yet are often used as shortcuts that require little effort. Most Academic stakeholders are heavily dependent on artificial intelligence tools for their daily activities. (Tan, Cheng, & Ling, 2025). As a result, various academic activities may hinder the development of essential skills such as self-improvement and critical thinking, and could reduce human expertise in collaborative AI environments (Klimova & Marcel, 2023). To address these challenges, further empirical studies are needed to clarify the appropriate roles of AI in learning, teaching, and assessment. This special issue seeks contributions that examine the impact of AI on these aspects of higher education (Sailer, et al., 2023).

The adoption of AI tools across various academic activities has significantly enhanced organisational processes not only in educational institutions but also in other

sectors (Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., et al., 2023). It covers various assessment strategies, student assistance services, and institutional governance in higher education (Luis & Cabanillas-García, 2025). Most studies have shown that AI tools have substantially enhanced the creation of learning and teaching content (Merk, Ophoff, & Kelava, 2023). Also, it has substantially increased content scalability and operational efficiency (Luis & Cabanillas-García, 2025). Despite the advantages AI tools have brought to academic institutions, they have also posted significant ethical challenges, as they have substantially compromised the integrity and reputation of academic Institutions (Cinar & Bilodeau, 2024). However, the integration of AI into the education sector has improved system performance in routine operations and across various academic activities. Also, findings remain segmented across disciplinary, regional, and application domains (Cinar & Bilodeau, 2024). According to Ward et al. (2024), AI tools significantly influence student study habits and academic performance by enabling automated feedback and personalized learning pathways. According to Merk et al (2025), most studies have shown that, from a security information science perspective, educational institutions operate as human socio-technical systems whose stability depends on confidentiality, integrity, availability, governance, and human agency, all of which are affected by AI integration (Mbah, Nugraha, & Kushnir, 2025). Terminology Note

In this study, the term "AI tools" is used as an umbrella term for artificial intelligence systems, applications, and technologies deployed in higher education contexts. This includes generative AI platforms, learning analytics systems, intelligent tutoring systems, and administrative AI applications.

1.2 Motivation for the Review

Artificial Intelligence is a technological tool designed to offer personalized and efficient methods to improve the delivery of educational instruction (Mbah, Nugraha, & Kushnir, 2025). According to the source, most studies have shown that AI tools are isolated, such as intelligent tutoring systems, automated grading, and learning analytics (BMJ, 2021). However, there is a limited systemic approach to how the adoption of AI tools has also empowered higher education ecosystems (UNESCO, 2026). Also, studies show that Existing reviews often focus on educational resources and learning activities that are increasingly disseminated to teachers and students through digital learning environments (Tan, Cheng, & Ling, 2025). It has also excluded recent advances in generative AI, thereby

limiting their explanatory and practical value to ecosystems (UNESCO, 2026). Nowadays, higher education institutions face great challenges related to academic integrity. Also, they face other changes such as research ethics, institutional autonomy in governance, and large-scale data management, which differ substantially from those encountered at lower levels of education (BMJ, 2021). Moreover, many studies have shown that the impact of AI in universities depends not only on technological capabilities but also on pedagogical activities and human agency (López-Zambrano et al., 2021). To ensure that Artificial Intelligence is used safely for all educational stakeholders, it is necessary to adhere to all ethical practices when implementing the technology in educational settings. To address these gaps, this review consolidates recent empirical and conceptual evidence on AI tools in higher education sectors published between 2020 and 2025 (Wang & Park, 2021). For their explicit attention to their benefits, challenges, and implications, thereby offering an integrated perspective to overall academic integrity, good practice, and governance (Tan, Cheng, & Ling, 2025).

1.3 Objectives and Research Questions

The objective of this systematic review is to assess the adoption of artificial intelligence tools in higher education systems from a security science perspective, focusing on their applications, institutional risks, governance implications, and cybersecurity challenges.

1.4 Theoretical Framework: Security Science Perspective

This study adopts a security science perspective to analyze the integration of artificial intelligence (AI) tools in higher education systems. Rather than viewing AI solely as an educational innovation, this framework conceptualizes AI as a socio-technical system that introduces institutional risks, security vulnerabilities, and governance challenges (European Union, 2023). The analysis is grounded in the Confidentiality, Integrity, and Availability (CIA) triad, a foundational model in information security. Confidentiality concerns arise from the extensive collection and processing of sensitive student and institutional data by artificial intelligence systems, often involving third-party platforms. Integrity risks include academic misconduct, content manipulation, algorithmic bias, and challenges in verifying authorship and originality in AI-assisted outputs. Availability refers to the reliability and continuous accessibility of AI-driven systems, particularly where institutions depend on external infrastructures or cloud-based services. In addition, the

study draws on socio-technical systems theory, which emphasizes the interaction among human actors, institutional processes, and technological systems (OECD, 2021). To operationalize security risks associated with AI tools in higher education, this study adopts a Weighted CIA Security Model defined as:

$$S\_AI = w\_c\ C + w\_i\ I + w\_a\ A$$

Where:

- S_AI represents the overall security risk score of an AI system.
- C (Confidentiality) measures the level of risk related to data exposure, privacy breaches, and unauthorized access to sensitive student and institutional data.

I (Integrity) reflect risks associated with academic misconduct, content manipulation, algorithmic bias, and loss of trust in AI-generated outputs. A (Availability) captures system reliability risks, including downtime, overdependence on third-party platforms, and service disruptions.

w_c, w_i, and w_a represent the weights assigned to each security dimension such that:

$$w\_c + w\_i + w\_a = 1$$

In higher education contexts, this study prioritizes integrity due to its direct impact on academic trust and assessment validity. Therefore, a typical weighting scheme is:

$$w\_c = 0.3,\ w\_i = 0.5,\ w\_a = 0.2$$

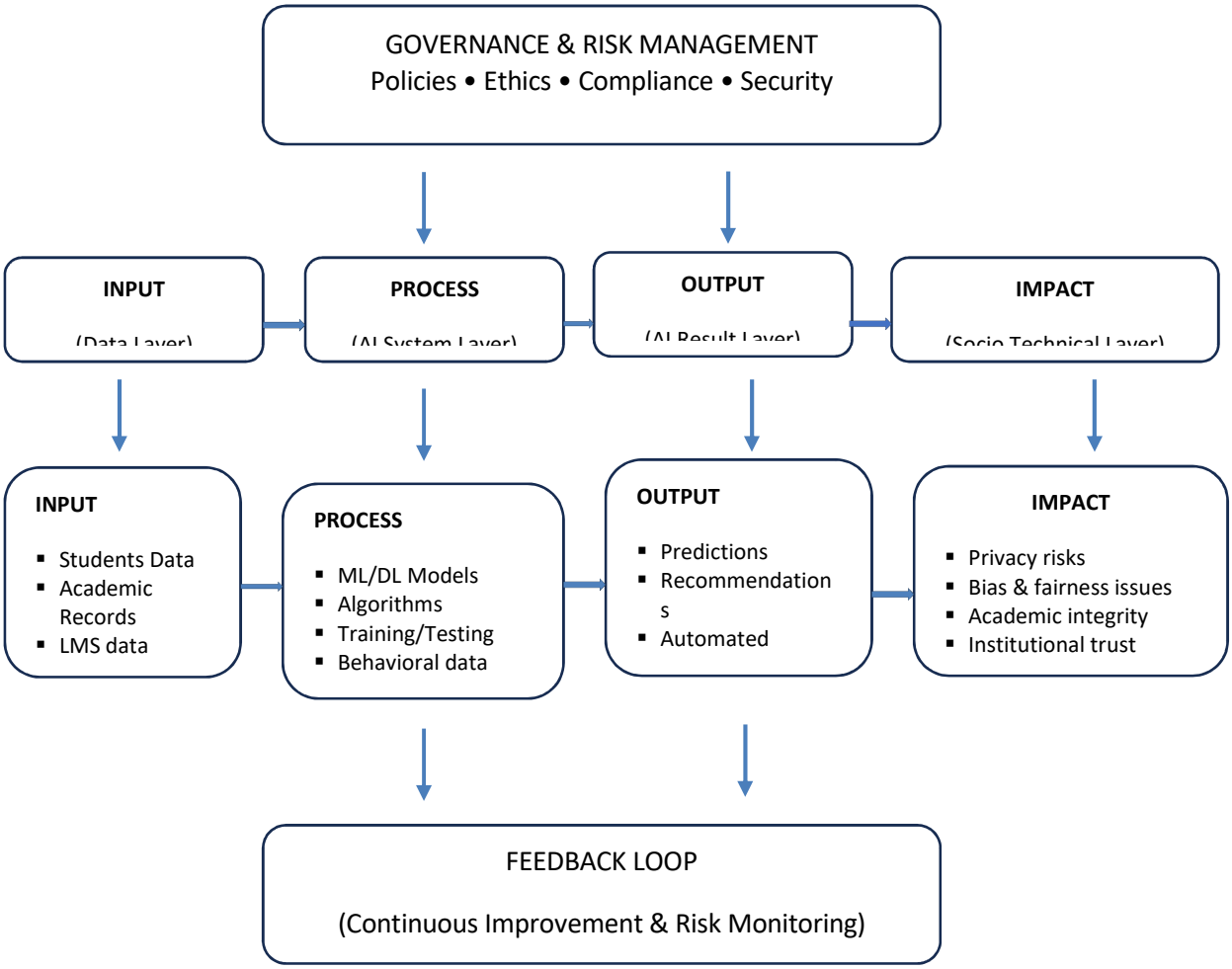
Each component (C, I, A) is operationalized on a normalized scale (0–1), where higher values indicate higher risk levels. The overall security score S_AI enables comparative assessment of different AI tools and supports risk-based governance decisions. In higher education, artificial intelligence tools operate not in isolation but as embedded components of institutional ecosystems, influencing teaching, assessment, governance, and decision-making. From a risk management perspective, the adoption of artificial intelligence introduces multiple layers of institutional risk, including data breaches, loss of academic integrity, overreliance on automated systems, and reduced human agency (ENISA, 2023). These risks require structured governance frameworks, including policy

development, accountability mechanisms, and continuous monitoring of the behavior of artificial intelligence systems. By integrating these security science concepts, this study reframes artificial intelligence adoption in higher education as a critical issue of institutional security, governance, and risk management, rather than as purely a pedagogical or technological advancement (Alzahrani, A., & Alghamdi, A., 2023).

1.5 Conceptual Framework

The framework conceptualizes artificial intelligence systems as interacting layers of input (data sources), processing (algorithms and models), output (AI-generated decisions and content), and impact (institutional, ethical, and security consequences), all embedded within governance and risk management structures.

Figure 1. Conceptual framework illustrating AI as a socio-technical security system in higher education.



Researcher 2026

The framework presents artificial intelligence systems as a layered socio-technical architecture consisting of four interconnected components: input, processing, output, and impact. The input layer comprises diverse educational data sources, including student records, learning management system data, and behavioral data. The processing layer involves machine learning and deep learning algorithms that transform input data into actionable insights. The output layer represents AI-generated predictions, recommendations, and automated decisions (Alzahrani, A., & Alghamdi, A., 2023).. The impact layer captures the broader institutional, ethical, and security implications, including privacy risks, bias, academic integrity concerns, and trust issues. These layers are governed by overarching governance and risk management structures, including policies, ethical guidelines, and regulatory frameworks. A feedback loop connects all components to enable continuous monitoring, evaluation, and improvement of AI system performance and security.

2. Methodology

2.1 Review Design

This study uses a systematic literature review approach to investigate the integration of artificial intelligence (AI) tools within academic systems (Ward, Bhati, Neha, & Guercio, 2024). While most recent research has focused primarily on applications, this review operationalizes a security science perspective, emphasizing institutional risk, good governance, and ethical responsibility for the sake of improving institutional integrity and reputations (Kahale, Piechotta, McKenzie, & Dorando, 2022). The review is guided by principles inspired by PRISMA 2020, aiming to address the complexity of artificial intelligence adoption across diverse academic activities (Assunção, Patrão, Castelo-Branco, & Menezes, 2022). The methodology extends beyond identifying AI technologies to examine the effects of their implementation on academic integrity, data governance, and institutional independence for all stakeholders in education(Tan, Cheng, & Ling, 2025).

Accordingly, a systematic review method was implemented, guided by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 statement to ensure transparency, rigor, and reproducibility (Assunção, Patrão, Castelo-Branco, & Menezes, 2022). The review is conducted in accordance with the PRISMA 2020 guidelines to promote the adoption of artificial intelligence tools across education systems. The protocol specified the objectives, eligibility criteria, search strategy, and screening

procedures prior to database searches. This review was not registered in a formal systematic review registry (Ward, Bhati, Neha, & Guercio, 2024).The review is further anchored in security science frameworks, particularly the CIA triad and socio-technical risk models, to systematically evaluate how artificial intelligence adoption affects institutional security, governance, and resilience.

2.2 Search Strategy

An extensive literature search was carried out across relevant databases. The review of this study was based on the following databases: Scopus, Web of Science, IEEE Xplore, ScienceDirect, SpringerLink, and Taylor & Francis. This selection is based on the relevance and coverage of these databases in AI and education research. It is significantly used to assess the real impacts and applications of Artificial Intelligence Tools (Assunção, Patrão, Castelo-Branco, & Menezes, 2022). The review included peer-reviewed publications from January 2020 to December 2025, a period marked by rapid adoption of artificial intelligence and the rise of generative artificial intelligence systems (Selwyn, 2022). Search strings combined Artificial Intelligence-related terms such as artificial intelligence, generative artificial intelligence, and machine learning. These terms were used to identify concepts related to higher education and security, such as academic integrity, data privacy, governance, and institutional risk (Assunção, Patrão, Castelo-Branco, & Menezes, 2022). This approach ensured coverage of both technological and governance-focused studies. Additional records were identified through ERIC and backward snowball sampling, as reflected in the PRISMA flow diagram (Selwyn, 2022). To enhance transparency and reproducibility, the search strategy employed Boolean search strings combining key concepts related to artificial intelligence, higher education, and security. An example of the search query used is as follows:

("Artificial Intelligence" OR "Generative AI" OR "Machine Learning")

AND ("Higher Education" OR "Universities")

AND ("Security" OR "Data Privacy" OR "Academic Integrity" OR "Governance")

These search strings were adapted slightly across databases to align with indexing requirements while maintaining conceptual consistency. To enhance transparency and reproducibility, database-specific search queries were adapted to align with indexing requirements. For example, in Scopus, the search was applied to titles, abstracts, and

keywords, while in Web of Science, topic-based searches were used. Filters were applied to include peer-reviewed articles published between 2020 and 2025 and written in English.

Table 1. Search Strategy and Keywords

Concept	Keywords
Artificial Intelligence	"Artificial Intelligence" OR "Generative AI" OR "Machine Learning"
Education	"Higher Education" OR "Universities"
Security	"Security" OR "Data Privacy" OR "Academic Integrity" OR "Governance"

The search strings were combined using Boolean operators (AND/OR) and adapted across databases.

2.3 Inclusion and Exclusion Criteria

This section introduces the overall process for screening the studies for inclusion or exclusion (Lameras & Arnab, 2022). The review critically examines the adoption of artificial intelligence tools in higher education institutions (Micheni, Machii, & Murumba, 2024). Eligible studies were also required to discuss broader implications, including benefits, risks, ethical concerns, or governance considerations associated with artificial intelligence adoption (BMJ, 2021). The review included only studies that focused on Artificial Intelligence in relation to Education systems to determine whether an article addressed the impact of teaching and learning (Micheni, Machii, & Murumba, 2024). The definition used during screening was taken from a meta-analysis of the impacts and applications of Artificial Intelligence (Tan, Cheng, & Ling, 2025). Alternatively, studies were excluded if they focused exclusively on primary or secondary education (Kenchakkanavar, 2023). Also, if they presented some of the technical aspects of artificial intelligence models without consideration, or if they lacked applicability to the educational system-level and governance-related context (Ward, Bhati, Neha, & Guercio, 2024). This filtering process ensured that the selected literature was analytically aligned with the objectives of security science and institutional governance analysis (Micheni, Machii, & Murumba, 2024). The study selection process followed a structured multi-stage screening procedure. First, all identified records were imported into a reference management system, and duplicate entries were removed. Second, title and abstract screening were conducted to exclude studies that did not meet the inclusion criteria. Third, full-text screening was performed to assess eligibility based on relevance to higher education and security-related

implications of artificial intelligence. Finally, the remaining studies were included for qualitative synthesis. The screening process was conducted by two independent reviewers to ensure reliability, with disagreements resolved through discussion and consensus.

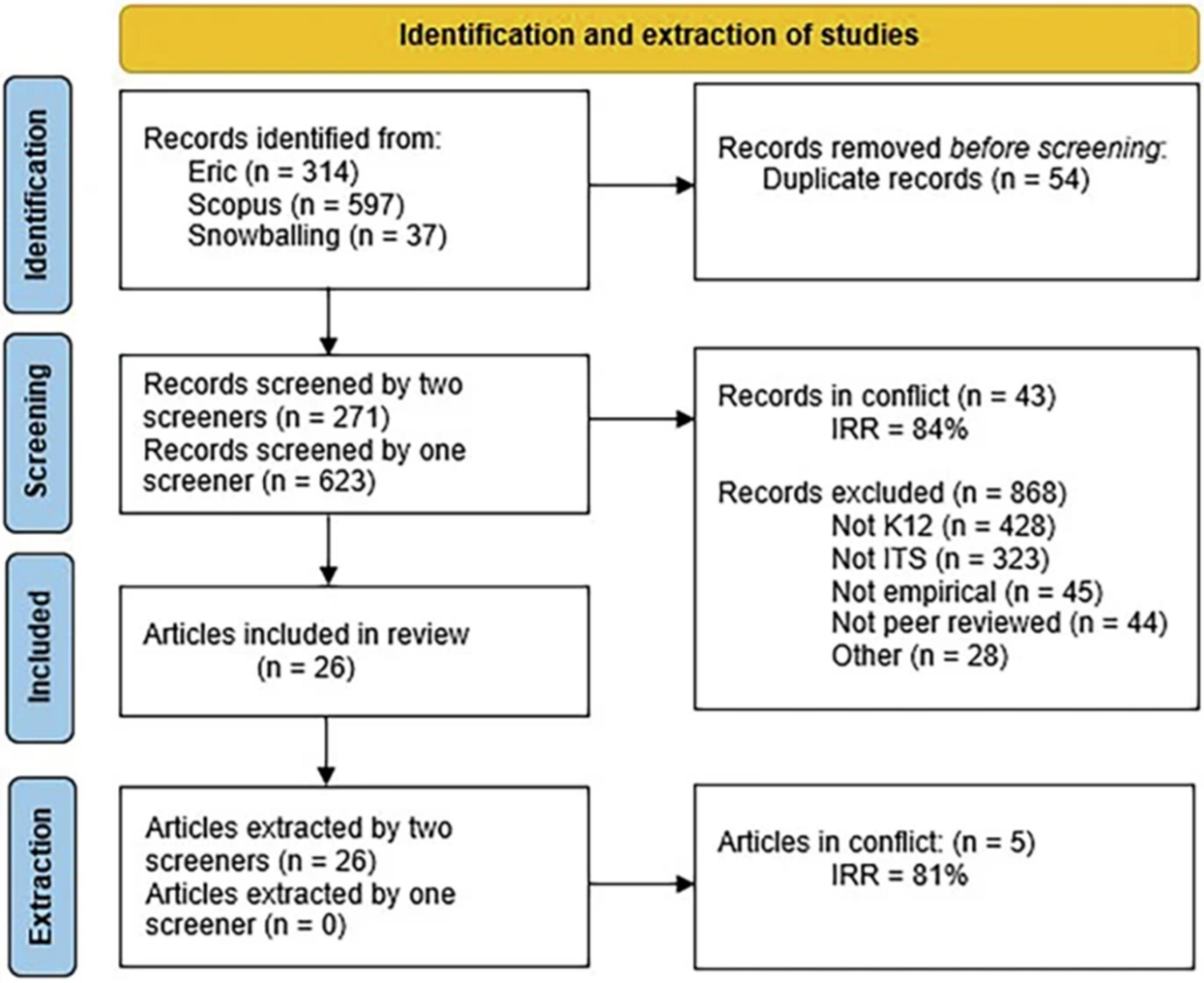
2.4 Data Extraction and Synthesis

Following titles and abstract screening, the study underwent full-text screening (Rizvi & Waite, 2023). It also shows that the general data extraction process was undertaken independently by two reviewers, with disagreements resolved through discussion and adjudication (Lameras & Arnab, 2022). This process involves identifying relevant information in each selected article and recording it in an extraction codebook within Covidence (Micheni, Machii, & Murumba, 2024). Data extracted from each study included the type of artificial intelligence tool used (e.g., generative artificial intelligence or learning analytics), the functional domain where the tool was applied, and the reported benefits of its use (Ward, Bhati, Neha, & Guercio, 2024). However, it describes the overall process of analysis, identifying risks, challenges, and governance, policy, and security implications arising from the adoption of artificial intelligence in academic Institutions (UNESCO, 2026). Based on the data extraction process, a thematic synthesis approach was employed to analyze the extracted data. Under these themes, refined iteratively around key issues such as academic integrity, data protection, human agency, and institutional governance (Tan, Cheng, & Ling, 2025). This method allowed artificial intelligence tools to be viewed as essential parts of institution-wide infrastructure, which were properly utilized rather than isolated or purely technical context.

2.5 Ethical Considerations

The review relied solely on secondary data from published sources. No human subjects were involved, so ethical approval was not necessary (Micheni, Machii, & Murumba, 2024). However, ethical issues discussed in the literature, especially related to misuse, surveillance, and accountability, were considered key analytical topics (Kenchakkanavar, 2023).

Figure 2. PRISMA 2020 flow diagram of the study selection process.



Source: Author's own elaboration based on PRISMA 2020 guidelines.

Figure 1 shows the PRISMA flow diagram, which outlines the process of identifying, screening, including, and extracting data from studies in this systematic review. The process: A total of 948 records were identified through database searches and additional methods, including ERIC (n = 314), Scopus (n = 597), and snowball sampling (n = 37). After the whole screening process, 54 duplicate records were removed (Micheni, Machii, & Murumba, 2024). During screening, more than 271 records were reviewed by two reviewers, while 623 were screened by one reviewer (Kenchakkanavar, 2023). The overall reliability (IRR) for the dual screening was 84%, with 43 records requiring conflict resolution. After title and abstract screening, more than 868 records were excluded for reasons such as not related to higher education institutions (K–12 focus) (n = 428), not involving artificial intelligence tools (n = 323), not empirical (n = 45), not peer-reviewed (n = 44), and other reasons (n = 28) (Ward, Bhati, Neha, & Guercio, 2024). Following full-text

review, 26 articles met the inclusion criteria and were included in the final sample. To ensure accurate results, these articles were independently extracted by two reviewers (n = 26), with none extracted by a single reviewer (Lameras & Arnab, 2022). The general reliability during data extraction was 81%, with 5 articles requiring adjudication to resolve disagreements. This thorough, multi-stage process ensured transparency, methodological rigor, and adherence to PRISMA guidelines for systematic reviews (Kenchakkanavar, 2023). The final inclusion of 26 studies reflects a rigorous filtering process based on predefined eligibility criteria, quality assessment, and relevance to the security-focused research objectives. This sample size is consistent with systematic reviews that prioritize depth of analysis and thematic synthesis over statistical generalization, ensuring that only the most relevant and high-quality studies were included.

2.6 Quality Assessment

To ensure the reliability and rigor of the included studies, a quality appraisal process was conducted using an adapted Critical Appraisal Skills Programme (CASP) checklist. Each study was evaluated based on criteria including clarity of research objectives, methodological rigor, validity of findings, relevance to the research questions, and consideration of limitations. Each criterion was scored on a three-point scale (1 = low quality, 2 = moderate quality, 3 = high quality). Studies scoring below a minimum threshold were excluded during full-text screening. This process ensured that only high-quality and methodologically sound studies were included in the final sample. The quality assessment further strengthened the validity of the synthesis and ensured alignment with systematic review best practices.

2.7 Risk of Bias Assessment

To assess methodological quality and potential bias, an adapted Critical Appraisal Skills Programme (CASP) checklist was applied. Each study was evaluated across five domains: clarity of objectives, methodological rigor, data validity, relevance to research questions, and reporting transparency.

Studies were scored on a three-point scale (low, moderate, high quality). Only studies meeting a minimum quality threshold were included in the final synthesis. This process minimized selection bias and enhanced the reliability of the findings.

3. Results: AI as a Socio-Technical Security System

This section presents the synthesized findings from the 26 studies included in the final analysis (Lameras & Arnab, 2022). The results are organized by the categories of artificial intelligence tools identified, their functional application areas within higher education systems, and their related governance and security implications (Rizvi & Waite, 2023).

3.1 Categories of AI Tools in Higher Education

The reviewed literature identifies four main categories of artificial intelligence tools used in higher education systems (Heron et al., 2023). Generative AI tools were the most prominent, discussed in more than half of the reviewed studies (n = 14) (Ward, Bhati, Neha, & Guercio, 2024). These include large language models and text-generation systems used for writing assistance, real-time feedback, content creation, and research support (Agormedah, Henaku, Ayite, & Ansah, 2020). Their rapid spread is due to their accessibility and ease of integration into existing higher education activities. Academic analytics and predictive systems comprised the second-largest category, appearing in about 9 studies (Smith, 2023). These systems use AI to evaluate student-related data to monitor engagement, predict academic performance, and identify at-risk students (Kenchakkanavar, 2023). They are often integrated into learning management systems and institutional systems. A third category comprises intelligent tutoring and adaptive learning systems, which were featured in seven studies (n = 7). These system categories offer institution-wide systems rather than individual applications, automated hints, and adaptive feedback, especially in STEM disciplines (Agormedah, Henaku, Ayite, & Ansah, 2020). Lastly, administrative and decision-support AI tools were discussed in fewer studies (n = 6) (Agormedah, Henaku, Ayite, & Ansah, 2020). This group includes chatbots for student services, real-time scheduling systems, and AI-supported enrollment and resource management tools (Smith, 2023).

3.2 Functional Areas of Application

Across the reviewed studies, Artificial Intelligence tools were applied in four main functional areas, with areas frequently overlapping within individual institutions. Teaching and learning applications were the most frequently reported recurring theme across the reviewed studies (n ≈ 20). These emphasized personalized learning, algorithm-driven feedback, content generation, and instructional planning (Tan, Cheng, & Ling, 2025).

Generative artificial intelligence emerged as a recurring theme both as a pedagogical support tool and a source of concern due to its impact on student learning process autonomy and improvement, critical thinking (Ward, Bhati, Neha, & Guercio, 2024). Assessment-related applications were discussed in around twelve studies (n = 12). These included AI-assisted grading, plagiarism detection, and feedback systems (Mamoon-Al-Bashir, Kabir, & Rahman, 2019). While efficiency improvements were widely noted, several studies raised concerns about academic integrity and authorship ambiguity. Administrative applications were reported in about 8 studies (n = 8), highlighting efficiency gains from automated student support, scheduling, and decision-making systems (Tan, Cheng, & Ling, 2025). These were often linked to issues of data privacy, surveillance, and dependence on external technology vendors (Ward, Bhati, Neha, & Guercio, 2024). Research-focused AI technologies were reported in a smaller but growing number of studies (n = 6), focusing on literature review support, and academic writing assistance, alongside concerns about research ethics, originality, and accountability(Mamoon-Al-Bashir, Kabir, & Rahman, 2019).

3.3. Governance, Risk, and Security Implications

According to Ward et al. (2025), several challenges arise from the adoption of artificial intelligence in educational institutions. These challenges, governance, and security considerations were explicitly addressed in only 5 studies (n = 5) (Aina & Abdulwasiu, 2023). Therefore, it was considered implicitly relevant in most applications (Mamoon-Al-Bashir, Kabir, & Rahman, 2019). From a security science standpoint, these studies highlighted that AI tools act as institution-wide socio-technical infrastructures rather than isolated educational technologies (Wang & Cheng, 2021). Artificial Intelligence tools have several consequences, including data leakage and algorithmic opacity (Pak, Polikoff, & García, 2020). Another challenge of integrating AI tools into academic institutions for educational purposes is that they can lead to academic misconduct and a decline in institutional reputation (Tan, Cheng, & Ling, 2025). Most existing studies reporting positive outcomes consistently pointed out the importance of formal governance frameworks, staff training, and clear accountability mechanisms (Wang & Cheng, 2021). Conversely, institutions lacking these structures were shown to be more susceptible to ethical breaches, policy inconsistencies, and security threats (Pak, Polikoff, & García, 2020).

4. Discussion

From a security science perspective, the findings demonstrate that artificial intelligence tools in higher education function as socio-technical systems that introduce both operational benefits and significant institutional security risks (Mamoon-Al-Bashir, Kabir, & Rahman, 2019). Artificial intelligence enhances educational technologies across academic integrity, institutional governance, data management, and decision-making, according to research. Despite these benefits, the integration of Artificial Intelligence introduces cybersecurity risks when used properly to enhance institutional efficiency (Tan, Cheng, & Ling, 2025). Also, sometimes Artificial Intelligence poses new security risks, including concerns about algorithmic transparency, data privacy, academic misconduct, and diminished human agency (Pak, Polikoff, & García, 2020). Issues with authorship ambiguity, difficulties in detecting plagiarism, and accountability in evaluations and research outputs have become more pressing due to the extensive use of generative AI. According to recent research, students and staff may become less capable of critical thinking, self-control, and epistemic responsibility if they rely too much on artificial intelligence (Aina & Abdulwasiu, 2023). One important conclusion is that good governance mechanisms significantly influence the effects of Artificial Intelligence innovation, not only in Academic sectors but also in other systems (Wang & Cheng, 2021). Institutions face heightened risks of ethical violations, data misuse, and loss of control if they lack robust policies, adequate staff training, or accountability. The widespread adoption of artificial intelligence within academic institutions should be viewed not merely as a proactive pedagogical innovation but also as a matter of institutional governance (Aina & Abdulwasiu, 2023). Their review further emphasizes the various tensions between the centralized control exercised by commercial Artificial Intelligence tools platforms and the institutional autonomy that underpins higher education systems (Kenchakkanavar, 2023). It also further highlights some issues related to data breaches and regulatory compliance that arise when relying on external artificial intelligence systems. particularly in areas where cybersecurity and data protection laws are evolving (Mamoon-Al-Bashir, Kabir, & Rahman, 2019). Effective artificial intelligence risk analysis requires careful procurement and continuous monitoring of AI system behavior from a risk management perspective (Pak, Polikoff, & García, 2020). All things considered, these results encourage a shift in AI governance from a technological advancement-focused to a system-oriented, human-centered, and security-conscious approach (Kenchakkanavar, 2023). The long-term viability

and reliability of higher education systems could be in jeopardy without this change (Tan, Cheng, & Ling, 2025). This study moves beyond descriptive synthesis by providing a structured security-oriented interpretation of AI adoption and proposing a governance model to guide institutional policy and risk management. Future research should empirically validate the proposed CIA-based risk model using real institutional datasets.

4.1 Limitations of the Review

Despite all the benefits that Artificial Intelligence brings to academic institutions, it also presents several limitations (Tan, Cheng, & Ling, 2025). First, the review focused primarily on Peer-reviewed scholarly works published in English from 2020 to 2025, potentially excluding relevant studies published in other languages (Martinez-Maldonado et al., 2022; Melzner et al., 2022). Second, although the review used a rigorous PRISMA-inspired methodology, it prioritizes thematic and system-level synthesis over formal risk-of-bias or quantitative effect-size assessments (Tan, Cheng, & Ling, 2025). Third, the changing nature of artificial intelligence technologies has highlighted that with the introduction of new tools, policies, and risks could emerge beyond the study period (Mamoon-Al-Bashir, Kabir, & Rahman, 2019). These limitations suggest that the findings should be viewed as a structured snapshot of current evidence rather than an exhaustive or static account of AI adoption in higher education (Pak, Polikoff, & García, 2020).

4.2 Comparative Risk Analysis of AI Tools

A comparative analysis of the reviewed studies shows that various categories of artificial intelligence tools pose different levels and types of security risks in higher education systems. Generative artificial intelligence tools stand out as the highest-risk category due to their direct role in content creation, which increases the risk of academic misconduct, authorship confusion, and misinformation. These tools notably challenge the integrity aspect of the CIA triad. Learning analytics and predictive systems primarily raise concerns about confidentiality, as they rely heavily on large volumes of sensitive student data. Key issues include data breaches, unauthorized access, and concerns about surveillance. Intelligent tutoring systems pose moderate risks, primarily from algorithmic bias and a lack of transparency, which impact both integrity and fairness in academic activities. Administrative AI systems, though they enhance efficiency, bring risks related to system dependency, availability, and third-party control. Overall, the analysis suggests that the level of risk is closely linked to the system's level of autonomy and data reliance.

4.3 Analytical Cross-Study Synthesis of AI Risk Patterns

A cross-study synthesis reveals consistent patterns linking the characteristics of artificial intelligence tools to specific security risks. Generative artificial intelligence systems pose the greatest integrity risk due to their autonomous content generation capabilities, which directly affect the authenticity of authorship and the validity of academic assessments. Their ability to produce human-like outputs increases the likelihood of undetectable academic misconduct. In contrast, learning analytics and predictive systems pose heightened confidentiality risks because they rely on large-scale aggregation of student data, often including sensitive behavioral and performance data. The risk is further amplified in institutions that lack strong data governance frameworks or rely on third-party platforms. Intelligent tutoring systems pose moderate integrity risks, primarily due to algorithmic bias and limited transparency in adaptive decision-making. These risks are more pronounced in contexts where explainability mechanisms are absent. Administrative AI systems exhibit relatively higher availability risks due to dependence on external infrastructures and cloud-based services. Institutions with limited technical capacity are particularly vulnerable to service disruptions and vendor lock-in. Across the reviewed studies, governance maturity emerges as a critical moderating factor. Institutions with formal artificial intelligence governance frameworks, staff training, and clear accountability structures consistently report lower overall risk levels. Conversely, weak governance environments amplify all three CIA risk dimensions, regardless of the AI tool category.

4.4 Critical Governance Gaps

The findings highlight several critical governance gaps that limit the secure and effective integration of artificial intelligence in higher education institutions. First, many institutions lack formal AI governance frameworks, resulting in inconsistent policies and weak accountability mechanisms. Second, there is limited institutional capacity for staff training and awareness of AI-related risks, which increases vulnerability to misuse and ethical violations. Third, reliance on third-party artificial intelligence platforms introduces challenges related to data ownership, compliance, and institutional autonomy. Many institutions lack clear guidelines for vendor risk assessment or data protection standards. Fourth, existing academic integrity policies are often outdated and do not adequately address AI-generated content. These governance gaps indicate that technological adoption

is outpacing institutional preparedness, creating a significant imbalance in security and risk management.

4.5 Illustrative Application of the Weighted CIA Model

Applying the model conceptually, generative AI tools typically exhibit high integrity risk ($I \approx 0.9$), moderate confidentiality risk ($C \approx 0.6$), and moderate availability risk ($A \approx 0.5$), resulting in a higher overall security score. In contrast, learning analytics systems pose a higher confidentiality risk ($C \approx 0.85$) but a lower integrity risk ($I \approx 0.6$), reflecting their data-driven nature. This demonstrates the model's usefulness in differentiating risk profiles across AI tool categories and guiding targeted governance strategies.

Table 2. Comparative CIA Risk Scores of AI Tools in Higher Education

AI Tool Category	Confidentiality (C)	Integrity (I)	Availability (A)	Weighted Score (S_AI)
Generative AI	0.6	0.9	0.5	0.72
Learning Analytics Systems	0.85	0.6	0.6	0.69
Intelligent Tutoring Systems	0.5	0.65	0.7	0.61
Administrative AI Systems	0.55	0.5	0.8	0.60

The results demonstrate that generative AI tools present the highest overall security risk, primarily due to their significant impact on academic integrity. In contrast, learning analytics systems exhibit elevated confidentiality risks due to extensive data usage. This quantitative comparison validates the applicability of the weighted CIA model for institutional risk assessment.

4.6 Proposed AI Risk Governance Model for Higher Education

Based on the findings of this study, a conceptual AI Risk Governance Model for Higher Education is proposed to address the identified security and governance challenges. The model integrates security science principles, socio-technical systems theory, and the AI risk lifecycle.

The model consists of four key layers:

1. Risk Identification Layer:

This layer focuses on identifying risks across the AI lifecycle, including data-related risks (input), algorithmic risks (processing), output-related risks (e.g., academic misconduct), and institutional impact risks.

2. Risk Assessment Layer:

This layer evaluates risks based on their likelihood and impact on confidentiality, integrity, and availability. It enables institutions to prioritize critical threats, such as data breaches and academic integrity violations.

3. Governance and Control Layer:

This layer includes policy frameworks, ethical guidelines, access control mechanisms, and accountability structures. It emphasizes the need for institutional oversight, regulatory compliance, and continuous monitoring of artificial intelligence systems.

4. Human-Centric Oversight Layer:

This layer ensures that human agency remains central to decision-making processes. It includes staff training, ethical awareness, and mechanisms for human intervention in AI-driven outcomes. The model operates within a continuous feedback loop, enabling institutions to monitor, evaluate, and adapt their AI governance strategies over time. This framework provides a structured approach for managing AI-related risks while supporting safe and sustainable adoption in higher education systems. The AI Risk Governance Model for Higher Education Systems presents a structured approach to managing risks associated with artificial intelligence. It begins with the Risk Identification Layer, where key risks such as data, algorithmic, output, and impact risks are recognized. These risks are then evaluated in the Risk Assessment Layer using the principles of confidentiality, integrity, and availability. This ensures that AI systems are secure, reliable, and accessible. At the core of the model is a continuous feedback and adaptation cycle that supports ongoing evaluation and improvement of artificial intelligence systems. This highlights that artificial intelligence governance is a dynamic and iterative process rather than a one-time activity. The model also emphasizes the importance of the Human-Centric Oversight Layer, which includes staff training, ethical oversight, and human intervention. These elements ensure that humans remain actively involved in monitoring and controlling AI systems. Overall, the framework integrates technical, organizational, and ethical considerations to promote responsible use of artificial intelligence.

5. Conclusion and Policy Implications

This review provides a comprehensive overview of artificial intelligence tools in higher education from 2020 to 2025, showing that AI adoption has advanced to a point where it directly impacts institutional security, governance integrity, and academic trust (Aina & Abdulwasiu, 2023). While AI offers significant benefits in teaching, assessment, research, and administration, its risks are equally considerable when implemented without strong oversight (Kenchakkanavar, 2023).

5.1 Key Contributions

This study offers multiple contributions to the existing literature and empirical review of artificial intelligence across higher education (Aina & Abdulwasiu, 2023). First, it offers a holistic system-level analysis that focuses solely on higher education institutions, moving beyond fragmented or application-specific analyses. Second, it combines pedagogical and governance perspectives with holistic, ethical perspectives on governance and security, offering a comprehensive understanding of AI adoption within institutional settings(Agormedah, Henaku, Ayite, & Ansah, 2020). Third, the study emphasizes generative artificial intelligence as a distinct and high-impact area with unique implications for academic integrity, governance, and risk management (Kenchakkanavar, 2023). Finally, it identifies AI adoption as a security-related issue rather than just a technical or instructional tool. Together, these contributions fill critical gaps in previous reviews that mainly examined the advantage of artificial intelligence from narrow technical or pedagogical viewpoints, thereby strengthening the evidence base for informed policy and institutional decisions (Pak, Polikoff, & García, 2020).

5.1.1 Primary Scientific Contribution

This study advances the existing literature by proposing a quantified artificial intelligence Risk Index for higher education systems, based on a weighted Confidentiality, Integrity, and Availability (CIA) model. Unlike prior studies that provide descriptive accounts of artificial intelligence adoption, this research operationalizes security risks into measurable components, enabling comparative evaluation of different artificial intelligence tools. The proposed model integrates three key dimensions, confidentiality, integrity, and availability, into a unified analytical framework, where each dimension is weighted according to its relevance in academic environments. This allows for systematic assessment of AI-related risks across teaching, assessment, administrative, and research contexts.

Furthermore, the study introduces a socio-technical artificial-intelligence Risk Governance Model that links risk identification, assessment, and control mechanisms to human-centric oversight. Together, these contributions provide a structured and scalable approach for evaluating and managing AI risks in higher education.

This dual contribution, combining a quantified risk model with a governance framework, offers both theoretical advancement and practical applicability, addressing a critical gap in research on AI in education.

5.2 Policy and Institutional Implications

Based on the findings of this review, the study addresses several cybersecurity and policy-relevant applications and implications of adopting artificial intelligence in the education sector (Agormedah, Henaku, Ayite, & Ansah, 2020). First, educational institutions should strengthen the governance of artificial intelligence frameworks to improve accountability and transparency, thereby promoting good artificial intelligence practices in academic activities. Also, they should ensure oversight across teaching, assessment, research, and administrative functions (Tan, Cheng, & Ling, 2025). Second, another advantage is that policies related to academic integrity and assessment security must be updated to address the use of generative AI, with an emphasis on learning design, attribution norms, and assessment strategies that minimize misuse rather than relying solely on detection technologies (Pak, Polikoff, & García, 2020). Third, Operationalization of AI technologies should proactively comply with cybersecurity standards (Wang & Cheng, 2021). This should be supported by clear protocols for data storage and protection, access control, and the management of third-party platforms (Agormedah, Henaku, Ayite, & Ansah, 2020). Fourth, safeguarding human agency requires continuous capacity building for both staff and students to promote responsible AI use, critical engagement, and awareness of ethical and security risks (Mamoon-Al-Bashir, Kabir, & Rahman, 2019). Finally, Educational organizations and other academic stakeholders should adopt a risk-aware approach to regulatory practices by conducting cybersecurity risk awareness training, risk-based evaluations of artificial intelligence vendors, and prioritizing explainability, interoperability, and compliance with applicable regulatory frameworks (Tan, Cheng, & Ling, 2025).

5.3 Directions for Future Research

Future research should move beyond descriptive accounts of artificial intelligence adoption to examine its broader and long-term implications for higher education systems (Wang & Cheng, 2021). In particular, scholars should examine the long-term security effects of artificial intelligence on institutional governance and compare regulatory strategies across different regional and legal contexts(Agormedah, Henaku, Ayite, & Ansah, 2020). Making artificial intelligence tool integrity and risk assessment frameworks adapted to the specific characteristics of higher education institutions (Pak, Polikoff, & García, 2020). Additionally, greater attention is needed to the human-interagency collaboration model that maintains academic independence and human agency. Enhancing research capacity in these areas would enhance the evidentiary support base for the safe, ethical, and resilient integration of artificial intelligence in higher education (Tan, Cheng, & Ling, 2025).

5.4 Final Remark

In conclusion, existing literature reviews have primarily focused on the operationalization of learning overview across various educational systems. Artificial intelligence tools constitute a strategic institutional capability and an operationalized risk for academic sectors (Mamoon-Al-Bashir, Kabir, & Rahman, 2019). The proper integration of artificial intelligence tools into academic systems depends not only on technological advancement but also on other features, such as strong governance, ethical safeguards, and security-conscious institutional design (Aina & Abdulwasiu, 2023). Considering AI as a security-relevant socio-technical system is crucial for protecting the integrity, autonomy, and sustainability of higher education in the digital age (Pak, Polikoff, & García, 2020).

Funding

This research received no external funding.

Conflicts of Interest

The author declares no conflict of interest.

Declaration of Originality

I declare that this manuscript is original, has not been published before, and is not under review elsewhere. All sources used in this study have been properly acknowledged, and the manuscript adheres to ethical standards for academic research.

References

Agormedah, E. K., Henaku, E. A., Ayite, D. K., & Ansah, A. E. (2020). Online learning in higher education during COVID-19 pandemic: A case of Ghana. Journal of Educational Technology and Online Learning. https://doi.org/10.31681/jetol.726441

Aina, K. J., & Abdulwasiu, A. A. (2023). Teachers' effective use of educational resources and their effect on students' learning. https://doi.org/10.22521/unibulletin.2023.122.4

Akmeşe, Ö. F., Kör, H., & Erbay, H. (2021). Use of machine learning techniques for the forecast of student achievement in higher education. Information Technologies and Learning Tools. https://journal.iitta.gov.ua/index.php/itlt/article/view/4178

Alzahrani, A., & Alghamdi, A. (2023). The role of artificial intelligence in enhancing education: Opportunities and challenges.

Assunção, G., Patrão, B., Castelo-Branco, M., & Menezes, P. (2022). An overview of emotion in artificial intelligence. IEEE.

Bahroun, Z., Anane, C., Ahmed, V., & Zacca, A. (2023). Transforming education: A comprehensive review of generative artificial intelligence in educational settings. Sustainability, 15(17), 12983. https://doi.org/10.3390/su151712983

BMJ. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. BMJ, 372, n71. https://doi.org/10.1136/bmj.n71

Chiu, T. K., Xia, Q., Zhou, X., Chai, C. S., & Cheng, M. (2023). Systematic literature review on opportunities, challenges, and future research recommendations of artificial intelligence in education. Computers and Education: Artificial Intelligence, 4, 100124. https://doi.org/10.1016/j.caeai.2023.100124

Cinar, A. B., & Bilodeau, S. (2024). Incorporating AI into the inner circle of emotional intelligence for sustainability. Sustainability. https://doi.org/10.3390/su16156648

Cordero, J., Torres-Zambrano, J., & Cordero-Castillo, A. (2024). Integration of generative artificial intelligence in higher education: Best practices. Education Sciences. https://doi.org/10.3390/educsci15010032

European Commission. (2019). Ethics guidelines for trustworthy AI. High-Level Expert Group on Artificial Intelligence. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

European Commission. (2023). Artificial Intelligence Act: Risk-based regulatory framework. https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence

European Union Agency for Cybersecurity (ENISA). (2023). Artificial intelligence threat landscape. https://www.enisa.europa.eu/publications/artificial-intelligence-threat-landscape

Heron, L., Buitrago-Garcia, D., Ipekci, A., Baumann, R., Imeri, H., & Salanti, G. (2023). How to update a living systematic review and keep it alive during a pandemic: A practical guide. Systematic Reviews. https://doi.org/10.1002/asi.24851

Kahale, L. A., Piechotta, V., McKenzie, J. E., & Dorando, E. (2022). Extension of the PRISMA 2020 statement for living systematic reviews. F1000Research. https://doi.org/10.12688/f1000research.75449.2

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., & Gurevych, I. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. Computers and Education: Artificial Intelligence.

Kenchakkanavar, A. Y. (2023). Exploring artificial intelligence tools: Realizing the advantages in education and research. https://doi.org/10.5281/zenodo.10251142

Klimova, B., & Pikhart, M. (2021). Exploring the effects of artificial intelligence on student and academic well-being in higher education: A mini-review. Frontiers in Psychology, 12, 635450. https://doi.org/10.3389/fpsyg.2021.635450

Lameras, P., & Arnab, S. (2022). Power to the teachers: An exploratory review on artificial intelligence in education. Information. https://doi.org/10.3390/info13010014

Luis, J., & Cabanillas-García, J. (2025). International trends and influencing factors in the integration of artificial intelligence in education. Education Sciences.

Mamoon-Al-Bashir, K., Kabir, R., & Rahman, I. (2019). The value and effectiveness of feedback in improving students' learning. Journal of Education and Practice.

Mbah, F. M., Nugraha, T. R., & Kushnir, I. (2025). Challenges and opportunities for leveraging generative AI for sustainability education: A critical review. Sustainability. https://doi.org/10.3390/su172310623

Merk, S., Ophoff, J. G., & Kelava, A. (2023). Rich data, poor information? Teachers' perceptions of graphical feedback. Learning and Instruction. https://doi.org/10.1016/j.learninstruc.2022.101717

Micheni, E. M., Machii, J., & Murumba, J. (2024). The role of artificial intelligence in education. Open Journal of Information Technology.

Noroozi, O., Khalil, M., & Banihashem, S. K. (2025). Artificial intelligence in higher education: Impact depends on support, pedagogy, human agency, and purpose. Interactive Learning Environments. https://doi.org/10.1080/14703297.2025.2539579

OECD. (2021). OECD framework for the classification of AI systems.

Pak, K., Polikoff, M. S., & García, E. S. (2020). The adaptive challenges of curriculum implementation. AERA Open. https://doi.org/10.1177/2332858420932828

Rizvi, S., & Waite, J. (2023). Artificial intelligence teaching and learning in K-12. Computers and Education: Artificial Intelligence.

Sailer, M., Bauer, E., Hofmann, R., Kiesewetter, J., Glas, J., & Gurevych, I. (2023). Adaptive feedback from artificial neural networks. Learning and Instruction. https://doi.org/10.1016/j.learninstruc.2022.101620

Selwyn, N. (2022). The future of AI and education: Some cautionary notes. European Journal of Education, 57(4), 620–631. https://doi.org/10.1111/ejed.12532

Smith, L. C. (2023). Reviews and reviewing: Approaches to research synthesis. Annual Review of Information Science and Technology. https://doi.org/10.1002/asi.24851

Tan, X., Cheng, G., & Ling, M. H. (2025). Artificial intelligence in teaching and teacher professional development: A systematic review. Computers and Education: Artificial Intelligence. https://doi.org/10.1016/j.caeai.2024.100355

UNESCO. (2026). Artificial intelligence in education. https://doi.org/10.54675/PCSP7350

Wang, T., & Cheng, E. (2021). Barriers to incorporating artificial intelligence in education. Computers and Education, 162, 104099. https://doi.org/10.1016/j.compedu.2020.104099

Ward, B., Bhati, D., Neha, F., & Guercio, A. (2024). Analyzing the impact of AI tools on student study habits and academic performance. arXiv. https://arxiv.org/abs/2412.02166