Artificial Intelligence in Higher Education: Ethical Risks and Transformative Potential. A Systematic Review

La Inteligencia Artificial en la universidad, riesgos éticos y potencial transformador: una revisión sistemática

Justina Limachi Manani
justylima@gmail.com
https://orcid.org/ 0009-0002-1820-328X
Universidad Pública de El Alto. La Paz, Bolivia

Hugo Alanoca Limachi
hugoalanocalimachi@gmail.com
https://orcid.org/0009-0000-4663-1631
Universidad Pública de El Alto. La Paz, Bolivia

ABSTRACT

Intelligent systems are transforming university education by personalising instruction and optimising academic management in innovative ways. This article aimed to describe the state of research on both the ethical challenges and transformative opportunities arising from the integration of Artificial Intelligence (AI) in higher education. A qualitative systematic review was conducted, selecting 34 articles published between 2021 and 2025, with a synthetic–interpretive qualitative approach. The findings reveal a dual landscape: on one side, AI enables personalised learning, accessibility, administrative efficiency, and predictive analytics for student retention; on the other, it poses significant ethical risks, including algorithmic bias, system opacity, threats to privacy, and the dehumanisation of the educational experience. In conclusion, realising AI's transformative potential requires a robust governance framework to ensure the technology serves the fundamental humanistic and educational purposes of higher education.

Keywords: Ethical challenges; Higher education; Governance; Artificial intelligence; Personalized learning.

RESUMEN

Los sistemas inteligentes transforman la formación universitaria al personalizar la instrucción y optimizar la gestión académica de forma novedosa. El propósito del presente artículo fue describir el estado de las investigaciones relacionadas con los desafíos éticos y las oportunidades de transformación que presenta la integración de la Inteligencia Artificial en la educación superior. La metodología utilizada fue una revisión sistemática cualitativa, se seleccionaron 34 artículos comprendidos entre 2021 y 2025, con un enfoque cualitativo de tipo sintético-interpretativo. Los resultados constatan que esta implementación presenta una dualidad: ofrece personalización del aprendizaje, accesibilidad, eficiencia administrativa y analítica predictiva para la retención estudiantil. Sin embargo, conlleva riesgos éticos profundos como sesgos algorítmicos, opacidad de los sistemas, amenazas a la privacidad y la deshumanización de la experiencia educativa. En conclusión, materializar su potencial transformador requiere un marco de gobernanza robusto que garantice que esta tecnología sirva a los fines humanistas y educativos fundamentales de la enseñanza superior.

Palabras clave: Desafíos éticos; Educación superior; Gobernanza; Inteligencia Artificial; Personalización del aprendizaje.

INTRODUCTION

Artificial Intelligence (AI) has burst onto the global scene as a disruptive force with the potential to reconfigure entire sectors, and higher education is no exception. In recent decades, universities have explored integrating digital technologies to improve their processes; however, AI represents an unprecedented qualitative leap. Its capacity to process big data, adapt in real time, and automate complex tasks promises to realise long-standing educational ideals, such as large-scale personalised learning, administrative efficiency, and equity in access to knowledge. This promise of a more responsive, inclusive, and efficient educational ecosystem places AI at the centre of the debate over the future of universities (Hutson et al., 2022).

The integration of AI in higher education results in a transformation toward a personalized and accessible learning experience. Through adaptive systems and intelligent platforms, AI can analyse individual student performance, pacing, and learning styles to provide tailored content and exercises. This approach enables students to progress at their own pace, address weaknesses, and deepen their strengths while reducing faculty workload for repetitive assessment tasks. Educators can consequently allocate more time to mentoring, critical discussion, and individual guidance, thereby enhancing classroom interaction quality (Gligorea et al., 2023).

Moreover, AI serves as a significant catalyst for administrative efficiency and academic research. Within administration, virtual assistants and chatbots facilitate responses to routine queries regarding enrollment, schedules, and regulations, thereby streamlining procedures and enhancing the student experience (Wongmahesak et al., 2025). In academic research, AI algorithms process large datasets rapidly, identify patterns, generate hypotheses, and expedite discoveries across disciplines, including medicine and the social sciences. This capability not only broadens students' knowledge but also equips them for a labour market where collaboration with intelligent systems is increasingly the norm (Kotsis, 2025).

Nonetheless, this transformative potential coexists with far-reaching ethical, technical, and pedagogical challenges. The integration of AI in university contexts is far from being a merely technical or instrumental issue. On the contrary, it implies a fundamental renegotiation of values central to the humanistic mission of higher education, such as equity, autonomy, privacy, academic integrity, and the very nature of the pedagogical relationship (Yadav, 2024). Emerging literature shows an inherent tension between the drive toward technological innovation and the urgent need to establish robust governance frameworks to prevent the risks of algorithmic bias, granular surveillance, dehumanization of the educational experience, and the erosion of human agency (Fowler, 2023).

66

Artificial Intelligence in Higher Education: Ethical Risks and Transformative Potential. A Systematic Review

This context presents specific features in regions such as Latin America. The adoption of AI in Latin American higher education faces particular structural challenges, including pronounced digital divides and socioeconomic disparities, insufficient technology and connectivity in many institutions, the requirement for local capacity building in AI research, development, and governance to prevent technological dependence, and the necessity for regulatory frameworks that protect educational community rights while supporting responsible innovation (Neri & Domingos, 2023).

Analysing how AI intersects with higher education is essential to ensure that the Fourth Industrial Revolution helps reduce, not reinforce, persistent inequalities in professional training. We must ask: What ethical challenges does integrating AI into higher education raise, and how do these affect core institutional values? How can AI truly transform and improve the university environment, beyond simple automation? This systematic review aims to summarise current research on the ethical challenges and transformative opportunities arising from the integration of Artificial Intelligence into higher education.

METHOD

This systematic review used a documentary search to identify, analyse, and synthesise academic literature on ethical criteria and challenges in applying AI to university education.

Search strategy and selection criteria

The search strategy was implemented across high-impact specialised databases and search engines, with emphasis on Scopus and Web of Science, to ensure access to high-quality international literature for the period 2021–2025. As a complement, university institutional repositories, Google Scholar, and specialized education journals were consulted to identify emerging research and diverse perspectives.

Queries combined the terms ethics, artificial intelligence, higher education, and university education to define the field precisely.

Scopus

The search was conducted in Scopus using the following query:

(TITLE-ABS-KEY ("artificial intelligence" OR "machine learning" OR "deep learning" OR "neural network*" OR "AI ethics") AND TITLE-ABS-KEY (ethic* OR "responsible ai" OR "algorithmic bias" OR "ai governance" OR fairness) AND TITLE-ABS-KEY ("higher education" OR "university education" OR "tertiary education" OR "college teaching" OR "post-secondary education"))

Filters applied:

- **Years:** 2021–2025 (to capture the most recent literature, given the rapid evolution of the field).
- **Document type:** Article, Review, Conference Paper, Book Chapter.
- **Subject area:** Social Sciences; Computer Science; Arts and Humanities.
- **Language:** English, Spanish (and others, depending on the scope of the review).

Web of Science

Search strategy

The following search query was used:

(TI = (("artificial intelligence" OR "machine learning" OR AI) AND (ethic* OR "responsible AI" OR fairness) AND ("higher education" OR "universit* education" OR "tertiary education"))) OR AB = (("artificial intelligence" OR "machine learning" OR AI) AND (ethic* OR "responsible AI" OR fairness) AND ("higher education" OR "universit* education" OR "tertiary education")) OR AK = (("artificial intelligence" OR "machine learning" OR AI) AND (ethic* OR "responsible AI" OR fairness) AND ("higher education" OR "universit* education" OR "tertiary education")))

Filters:

- **Years:** 2021–2025.
- **Document type:** Article, Review, Proceedings Paper, Book Chapter.
- **Web of Science categories:** Education & Educational Research; Ethics; Computer Science, Artificial Intelligence; Social Sciences, Interdisciplinary. To ensure the robustness and validity of the review, rigorous selection criteria were established. Priority was given to recent publications, mainly from the last five years, articles indexed in Q1 and Q2 journals, and academic books from well-established scholarly publishers. This approach made it possible to focus the analysis on high-impact and up-to-date contributions, which is essential in a field as dynamic and rapidly evolving as artificial intelligence. Thematic relevance and methodological quality guided the selection process, and sources that, although related, lacked sufficient analytical depth or empirical rigor were excluded.

Information retrieval and management process

The process was optimized through the use of the Zotero reference manager, which was essential for efficiently organizing references, storing full-text documents, and ensuring consistent citation throughout the study. Finally, the quality assessment of the selected sources was not limited to impact indicators but also considered clarity of presentation, relevance, and internal coherence. This comprehensive evaluation ensured that the present review is based on a robust, reliable, and up-to-date documentary foundation, providing a solid panoramic and critical overview of the intersection between artificial intelligence and ethics in the university context.

RESULTS

The search and selection process, detailed in the methodological section, enabled the identification and analysis of a final documentary corpus consisting of 34 publications, which constitutes the empirical basis for the findings presented below. The results of this review are organized around two main analytical dimensions that emerged from the examined literature: first, the ethical challenges and risks identified in the implementation of AI in university contexts are presented; second, the opportunity horizons are

68

Artificial Intelligence in Higher Education: Ethical Risks and Transformative Potential. A Systematic Review

outlined, highlighting the transformative potential of AI in higher education.

This triangulation of perspectives—focused on the problem, the foundations, and the solutions—offers a comprehensive and critical overview of the state of the art, revealing both emerging areas of consensus and significant issues that remain under debate and require further research.

Main ethical challenges in integrating artificial intelligence in higher education

The integration of AI into higher education ecosystems promises to transform pedagogical and administrative practices through learning personalization, process automation, and predictive analytics. However, this emerging paradigm entails complex and systemic ethical challenges that, if left unmitigated, may compromise fundamental values such as equity, autonomy, and academic integrity. The literature analysis reveals a consensus on six critical areas of concern that are not isolated but frequently appear in interconnected, synergistic ways.

First, data privacy and security are primary concerns. According to Amén et al. (2024), AI systems operate by collecting and processing large-scale student data (Big Data), which may include academic performance, as well as behavioral and interaction patterns. This procedure, as Campos et al. (2025) argue, creates a latent risk of granular surveillance and calls into question the nature of informed consent, which is often obtained through opaque terms of use that users accept tacitly. The security of these sensitive data repositories against cyberattacks constitutes another critical facet of this challenge.

Second, the promise of equity clashes directly with the risk of algorithmic bias and systemic discrimination. Algorithms are not neutral entities; they reflect the historical or social data on which they are trained. If these data contain existing prejudices—for example, related to gender, ethnicity, or socioeconomic status—the system may perpetuate and even amplify these inequalities, leading to discriminatory decisions in assessment, vocational guidance, or scholarship allocation, as Selgas (2025) argues.

This inherent bias is further aggravated, according to Bustelo (2025), by the opacity of black-box models, which constitutes the third challenge: a lack of transparency and explainability. Thus, Tweedie and Sharmi (2025) consider that when neither educators nor students can understand the logic behind an automated decision, accountability and the possibility of effective recourse are undermined, eroding trust in the institution.

A fourth problematic axis concerns the threat to academic integrity and authorship. The emergence of generative AI has blurred the boundaries of original work, raising unprecedented dilemmas regarding authorship, intellectual property, and the assessment of genuine learning. For Yeo (2023), Vasylyshyna et al. (2024), and Wise et al. (2024), reliance on these tools without precise regulation can foster sophisticated plagiarism and compromise the acquisition of fundamental competencies.

These technical and academic challenges converge into two broader problems. On the one hand, Ahmed (2024) addresses equity and access, where the risk is not only algorithmic but also material. The digital divide may exacerbate existing inequalities if universal access to technology and the digital literacy needed to use it critically are not ensured. On the other hand, according to Srishti (2024) and Dang and Liu (2025), there is the risk of dehumanizing the educational experience. Excessive reliance on AI-mediated interactions could devalue the irreplaceable role of teachers as mentors, guides, and role models, as well as erode spaces for dialogue and collaborative knowledge construction that are essential for the development of socioemotional skills and critical thinking, as Gao (2025) notes. This final point goes beyond operational efficiency and speaks to the very core of the university's humanistic mission.

In summary, for the authors cited above, the ethical challenges of AI in higher education are not mere technical obstacles, but deeply pedagogical, social, and philosophical issues that require a robust governance framework, forward-looking regulation, and continuous reflection by the entire university community.

Opportunity horizons: the transformative potential of AI in higher education

In the face of this necessary and critical landscape of ethical challenges, AI simultaneously emerges as a first-order catalyst with significant potential to enrich and structurally reconfigure university ecosystems.

1. Moving beyond technological determinism

Far from being mere automation tools or technocratic instruments, AI applications—when implemented with solid pedagogical grounding and a strategic vision—can redefine pedagogical, administrative, and inclusion practices, bringing the university institution closer to the ideals of personalization, efficiency, and equity that have guided its evolution for decades. As George (2023) and Salma and Ahmed (2024) note, the actual value of AI in education does not lie in replacing educators, but in amplifying their capabilities and freeing them from routine tasks, enabling them to focus on what is intrinsically human: mentoring, inspiration, and the facilitation of profound learning experiences.

The analysis conducted by Strielkowski et al. (2025) identifies and develops five interdependent key opportunity areas, where the fundamental added value lies in AI's capacity to offer scalable, highly adaptive, and data-driven solutions that transcend the limitations of traditional educational models.

2. Learning personalization at scale: from mass instruction to fine-tuned adaptation

AI's most paradigmatic contribution is, according to Laak and Aru (2025), its capacity to realize learning personalization at scale—an objective long pursued but technically unattainable in mass environments. This concept goes far beyond allowing students to progress at their own pace. For Halkiopoulos and Gkintoni (2024) and de Mooij et al. (2025),

70

Artificial Intelligence in Higher Education: Ethical Risks and Transformative Potential. A Systematic Review

it is a dynamic and recursive process in which AI systems, through machine learning algorithms and educational data mining, can continuously diagnose prior knowledge, cognitive learning styles, metacognitive states, and even each student's affective–motivational factors.

From a technological standpoint, Lin et al. (2023) argue that this is embodied in the development of Intelligent Tutoring Systems (ITS) and adaptive learning environments. An ITS, as explained by Singh (2024) and Singh et al. (2025), integrates four core modules: the domain model (expert knowledge), the student model (a dynamic representation of the learner's knowledge), the pedagogical model (instructional strategies), and the user interface. As the learner interacts with the system, the ITS continuously updates its model, identifying knowledge gaps and specific misconceptions. In response, it not only recommends specific resources (from an explanatory video to an academic article), but also adjusts the sequence and difficulty of problems, provides personalized cognitive scaffolding, and generates examples contextualized to the learner's interests.

Empirical evidence is beginning to support this potential. Basri (2024) in the field of exact sciences has shown that students who use ITS achieve significantly greater learning gains than those in traditional environments, mastering the subject matter in less time. This represents a break from the one-size-fits-all mass instruction model, moving toward an ecosystem in which education adapts to the individual rather than the other way around.

3. Enhanced formative feedback and assessment: assessment as a process, not a product

Inextricably linked to personalization is the opportunity for enhanced formative feedback and assessment. Traditional assessment, primarily focused on final products and scalable only through substantial teaching effort, often arrives too late to be truly formative. For Vetrivel et al. (2025), AI is transforming this paradigm, turning assessment into a continuous, integrated, and enriching process.

According to Wilson et al. (2022), AI systems equipped with Natural Language Processing (NLP) can analyze open-ended responses, essays, and written compositions, evaluating not only grammatical and spelling accuracy but also argumentative coherence, text structure, strength of ideas, and use of evidence. Platforms such as Turnitin or Grammarly, in their educational versions, already offer this type of preliminary feedback (Laflen, 2023). However, Banawan et al. (2023) argue that the horizon is more ambitious: systems such as Writing Pal provide interactive tutoring during the writing process itself, guiding students through planning, drafting, and revision.

Beyond writing, AI enables the automated assessment of oral presentations by analyzing nonverbal language, tone of voice, and clarity of speech. According to Feng et al. (2021), in science, technology, engineering, and mathematics disciplines, ITS can evaluate problem-solving step by step, identifying the exact moment a learner applies an incorrect

strategy and offering a conceptually appropriate hint. This not only reduces teachers' assessment workload—freeing up to 30% of their time—but also democratizes access to high-quality, immediate feedback, turning each task interaction into an opportunity for iterative and self-regulated learning.

4. Inclusion and accessibility: toward Universal Design for Learning

In the area of inclusion and accessibility, AI is positioned as a powerful tool for operationalizing the principles of Universal Design for Learning, which aim to create learning environments accessible to all from the outset, without the need for subsequent adaptations (Saborío & Rojas, 2024).

The AI-assisted tool repertoire is vast and has an immediate impact. For students with visual impairments, state-of-the-art text-to-speech systems powered by neural voice models offer unprecedented naturalness and fluency. Conversely, as Hyatt and Owenz (2024) indicate, high-accuracy speech recognition enables students with motor disabilities or dyslexia to interact with content and demonstrate knowledge through voice. Real-time machine translation and automatic caption generation not only benefit deaf or hard-of-hearing students but also international students and speakers of minority languages, breaking down language barriers.

5. Automating academic management: freeing human capital for value-added work

From an institutional perspective, automating academic management with AI represents tangible gains in operational efficiency and a revaluation of human capital. According to George and Wooden (2023), a significant portion of faculty and administrative staff time is consumed by repetitive, low-value-added tasks, generating professional burnout and diverting resources from the university's central mission.

These authors argue that intelligent chatbots can autonomously manage up to 80% of recurrent administrative inquiries (timetables, enrollment procedures, degree requirements), operating 24/7 and in multiple languages. Optimization algorithms can generate academic schedules that maximize space utilization, minimize student conflicts, and respect faculty preferences—an NP-hard problem that is unmanageable manually for large institutions.

Predictive AI systems for resource management can anticipate course demand, optimize scholarship and financial aid allocation, and forecast infrastructure needs. This comprehensive automation does not seek to eliminate jobs, but, as Koukaras et al. (2025) argue, to redefine them. By freeing professionals from routine tasks, institutions can redirect human talent toward higher-value activities, such as more personalized tutoring, the design of innovative learning experiences, high-quality research, and societal engagement. This is a human-centered optimization.

6. Predictive analytics for student retention: addressing dropout proactively

Finally, learning analytics strengthened by AI-based predictive models emerges, according to Siddiqui et al. (2025), as a crucial tool for

72

Artificial Intelligence in Higher Education: Ethical Risks and Transformative Potential. A Systematic Review

student retention, a key indicator of quality and equity in any higher education system. University dropout is a multifactorial phenomenon, and AI enables a shift from a reactive model—intervening only after students have already failed or withdrawn—to a proactive, preventive one.

According to Airaj (2024), the most advanced predictive models do not focus solely on grades. They use multimodal data: patterns of access to the virtual platform (frequency, duration, resources consulted), forum participation, attendance history, and even—always under the strictest ethical safeguards—interaction data from online tutoring systems. Using techniques such as random forests or gradient boosting, these models identify complex, nonlinear patterns undetectable to the human eye and assign each student a probability of dropout risk. When this risk exceeds a predefined threshold, the system generates an early warning that is sent to guidance services, tutors, or the student. This enables timely and targeted interventions, such as personalized tutoring, supplementary academic support, or psychosocial counseling.

7. A potential conditioned by ethical governance

In sum, AI's transformative potential lies in its intrinsic capacity to contribute to the creation of a more adaptable, responsive, inclusive, and efficient educational ecosystem. From personalization to dropout prevention, its applications promise a qualitative leap in how higher education is conceived and managed. However, it is imperative to conclude by emphasizing that the effective and sustainable realization of these opportunities is not automatic and depends on technology.

It depends on a critical, contextualized implementation guided by a robust governance framework that anticipates and actively mitigates the ethical risks previously analyzed in the studies included in the review: privacy, bias, transparency, and dehumanization. The future is not AI replacing the university, but the university integrating AI wisely and humanely, ensuring that this powerful technology is unequivocally at the service of the educational and human purposes that constitute its core mission.

DISCUSSION

The integration of AI in higher education has generated a broad spectrum of research addressing both its ethical challenges and its transformative opportunities. The following presents a comparative analysis of this review's findings and the views of other relevant authors in the field, organized around the two central themes: ethical challenges and horizons of opportunity. This exercise enables the identification of areas of convergence, nuance, and significant contradiction.

The results of this review highlight privacy and data security as primary concerns, noting that AI systems collect students' big data, which poses risks of granular surveillance and questions the validity of informed consent (Amén et al., 2024; Campos et al., 2025). This stance aligns fully with Andreotta et al. (2022), who warn of the extractivist logic of personal

data in advanced technological systems, where consent becomes diluted in unintelligible terms of use.

Likewise, there is agreement with Chen et al. (2023), who, from an educational perspective, highlight data governance and power asymmetries between institutions and students. However, while this review emphasizes cybersecurity risks, Ferhataj et al. (2025) broaden the focus to the long-term implications for students' moral autonomy—an aspect only tangentially addressed here through the discussion of consent.

The finding that algorithms can perpetuate and amplify pre-existing social inequalities is a strong point of consensus in the literature. In this regard, the review converges with Williamson et al. (2023), who argue that the datafication of education is not neutral and reflects the social contexts from which data are drawn; processes, behaviors, and interactions that were not previously quantified or considered as data are transformed into digital traces to be measured, stored, analyzed, and monetized.

However, a subtle contradiction arises with more techno-optimistic perspectives, such as those of Chinta et al. (2024), who—while acknowledging the risk of bias—argue that predictive analytics can be designed to identify and correct inequities, thereby enhancing educational justice. This review, by contrast, aligns with a more cautious stance, emphasizing the opacity of black-box models, as Bustelo (2025) notes regarding the ethics of algorithms and their lack of accountability.

The review identifies the threat to academic integrity as a central problem area, particularly with the advent of generative AI (Yeo, 2023; Vasylyshyna et al., 2024). This concern is widely supported by emerging literature. For example, Kotsis (2024) explores in depth how AI is redefining plagiarism and blurring the boundaries of authorship, in complete agreement with the findings presented here. However, while this review focuses on the risks of sophisticated plagiarism, Skoryk et al. (2025) introduce a contrasting nuance: they propose that, instead of banning these tools, higher education should integrate them to teach hybrid authorship and develop new forms of process-based assessment— a horizon that this review does glimpse in mentioning genuine learning assessment, although it does not fully develop the alternative pedagogical proposal.

The review presents large-scale personalized learning as AI's most paradigmatic contribution, implemented through ITS (Laak & Aru, 2025; Singh, 2024). This vision aligns with Baumgart and Madany (2022), who demonstrated that ITS effectiveness enhances knowledge acquisition in specific domains. It also coincides with the notion of adaptive learning as a means of closing performance gaps.

The potential of AI to transform assessment into a continuous, formative process (Vetrivel et al., 2025) aligns with Vashishth et al. (2024), who propose integrating assessment seamlessly into learning activities to provide immediate feedback.

74

Artificial Intelligence in Higher Education: Ethical Risks and Transformative Potential. A Systematic Review

Identifying AI as a tool for realising UDL (Saborío & Rojas, 2024) aligns with research on accessible technologies. As in this review, Ahmed et al. (2025) list multiple AI tools—such as speech synthesis and speech recognition—that remove barriers for students with disabilities. This perspective is broadly shared. However, a significant contradiction comes from Taylor et al. (2022), who question technological solutionism in inclusive education, arguing that the mere provision of tools does not guarantee inclusion unless accompanied by profound changes in institutional culture and practice. While this review adopts an optimistic tone regarding the arsenal of tools, it does not sufficiently problematize the institutional conditions and teacher training required for this promise to materialize.

The use of predictive analytics for proactive student retention, as proposed by Siddiqui et al. (2025) and Airaj (2024), is one of the findings most in line with mainstream research in learning analytics. The description of models for multimodal data is consistent with the frameworks proposed by Jiang et al. (2024). The central tension, however, lies in the ethics of intervention. Although this review mentions ethical safeguards, it does not delve into critiques of data paternalism, where interventions based on algorithmic predictions may limit student autonomy and create self-fulfilling prophecies. This critical perspective constitutes an essential counterpoint to the predominantly technical and optimistic view presented here on this specific topic.

In synthesis, the findings of this systematic review align closely with the main currents of the international literature on the identification of ethical risks, particularly regarding privacy, bias, and dehumanization. Its positions are framed within a critical and cautious perspective that has gained prominence in recent years. As for opportunities, the review aligns with the body of research that evidences AI's benefits for personalization, assessment, and inclusion, but tends to present a more harmonious and less conflictive picture than that emerging from some of the field's more critical voices.

The main contradictions lie not in the identification of phenomena themselves, but in the assessment of their depth, the nuances of their pedagogical implications, and the degree of centrality given to structural critique versus technological optimism. This analysis confirms that the state of the art on AI in higher education is a field of active dialogue, where consensus on problems coexists with open debates about solutions and their ultimate consequences.

CONCLUSIONS

The analysis of the studies included in the review shows that integrating AI into higher education represents a turning point with significant transformative potential. However, it also entails profound ethical challenges that require robust governance and continuous reflection. The following conclusions are derived from this analysis.

The implementation of AI systems in university settings depends on processing large volumes of student data, raising serious concerns about privacy and security. Informed consent, often obtained through opaque terms of use, is insufficient to guarantee user autonomy. Moreover, the vulnerability of such repositories to cyberattacks demands reinforced protection measures. The advantage of personalized learning thus collides with a latent risk of granular surveillance and loss of control over personal information.

One of the most significant ethical risks identified is the perpetuation and amplification of historical and social biases through algorithms trained on discriminatory data. This can affect critical decisions related to assessment, career guidance, or scholarship allocation, exacerbating pre-existing inequalities. Added to this is the digital divide, which limits equitable access to AI tools. However, when implemented in line with the principles of algorithmic justice, these same systems can foster more inclusive education tailored to student diversity.

The opacity of black-box models undermines accountability and trust in educational institutions. When neither teachers nor students understand the logic behind an automated decision, the possibility of appeal is limited, and system legitimacy is weakened. Transparency is, therefore, not only a technical requirement but a necessary condition for the ethical acceptance of AI in educational contexts.

The emergence of generative AI has blurred the boundaries of originality, raising unprecedented dilemmas about plagiarism, intellectual authorship, and the assessment of genuine learning. While these tools can facilitate sophisticated plagiarism, they also offer opportunities to redesign assessment processes, orienting them toward formative feedback and the development of metacognitive competencies.

A fundamental risk is the dehumanization of the educational experience, where excessive reliance on AI-mediated interactions may erode the pedagogical bond and collaborative knowledge construction. Nevertheless, when used in a complementary way, AI's potential lies in freeing teachers from routine tasks, allowing them to focus on mentoring, inspiration, and the development of socioemotional and critical skills.

Among the most notable advantages is AI's capacity to deliver personalized learning at scale through intelligent tutoring systems and adaptive environments that respond to each student's needs. Similarly, its application within the UDL framework facilitates accessibility for people with disabilities and students from diverse linguistic backgrounds, contributing to more equitable education.

The automation of administrative and academic processes through AI yields significant gains in operational efficiency while revaluing human capital by relieving it of repetitive tasks. Furthermore, learning analytics and predictive models enable early interventions to prevent student dropout,

76

Artificial Intelligence in Higher Education: Ethical Risks and Transformative Potential. A Systematic Review

provided they are used with ethical safeguards that protect privacy and autonomy.

In short, the future of AI in higher education will not depend solely on its technical capabilities but on the construction of governance frameworks that balance innovation and ethical responsibility. The sustainable realisation of its benefits—personalisation, inclusion, efficiency—is subject to a critical, context-specific implementation that mitigates risks such as algorithmic discrimination, opacity, and the erosion of humanistic values. AI should not replace the university, but be integrated into it in the service of its fundamental educational and social purposes.

REFERENCES

Ahmed, F. (2024). The digital divide and AI in education: Addressing equity and accessibility. AI EDIFY Journal, 1(2), 12-23. https://researchcorridor.org/index.php/aiej/article/view/259/248

Ahmed, S., Rahman, M. S., Kaiser, M. S. y Hosen, A. S. (2025). Advancing personalized and inclusive education for students with disability through artificial intelligence: perspectives, challenges, and opportunities. Digital, 5(2), 11. https://doi.org/10.3390/digital5020011

Airaj, M. (2024). Ethical artificial intelligence for teaching-learning in higher education. Education Information Technologies, 29(13), 17145-17167. https://doi.org/10.1007/s10639-024-12545-x

Amén, P., Zavala, D. L., Moran, N. S. y Intriago, A. B. (2024). Desafíos éticos y de privacidad en la implementación de la inteligencia artificial en la educación superior. Revista Científica Arbitrada de Investigación en Comunicación, Marketing y Empresa REICOMUNICAR, 7(14), 613-628. https://doi.org/10.46296/rc.v7i14.0286

Andreotta, A. J., Kirkham, N. y Rizzi, M. (2022). AI, big data, and the future of consent. Ai Society, 37(4), 1715-1728. https://doi.org/10.1007/s00146-021-01262-5

Banawan, M., Butterfuss, R., Taylor, K. S., Christhilf, K., Hsu, C., O'Loughlin, C., . . . McNamara, D. S. (2023). The future of intelligent tutoring systems for writing. In Digital Writing Technologies in Higher Education: Theory, Research, and Practice (pp. 365-383). Springer. https://doi.org/10.1007/978-3-031-36033-6_23

Basri, W. S. (2024). Effectiveness of AI-powered Tutoring Systems in Enhancing Learning Outcomes. Eurasian Journal of Educational Research(110). https://doi.org/10.14689/ejer.2024.110.003

Baumgart, A. y Madany, A. (2022). A knowledge-model for ai-driven tutoring systems. In Information Modelling and Knowledge Bases XXXIII (pp. 1-18). IOS Press. https://doi.org/10.3233/FAIA210474

Bustelo, J. L. (2025). Hacia una gobernanza algorítmica transparente: auditoría de sesgo. Estudio de caso. Cuadernos de Gobierno y Administración Pública, 12(1), e97604. https://dx.doi.org/10.5209/cgap.97604

Campos, V. A., Bastidas, K. A., Bastidas, D. L. y Alvarado, S. M. (2025). Fortalecimiento de la seguridad cibernética en universidad mediante inteligencia artificial para la detección de amenazas. Revista Social Fronteriza, 5(1). https://doi.org/10.59814/resofro.2025.5(1)607

Chen, J., Chen, Y., Cheng, X., Liu, J., Wang, Y. y Li, Y. (2023). Optimization Study of Higher Education Data Governance in the Era of AI. Proceedings of the 2023 International Conference on Information Education and Artificial Intelligence, https://doi.org/10.1145/3660043.36601

Chinta, S. V., Wang, Z., Yin, Z., Hoang, N., Gonzalez, M., Quy, T. L. y Zhang, W. (2024). FairAIED: Navigating fairness, bias, and ethics in educational AI applications. arXiv. https://doi.org/10.48550/arXiv.2407.18745

Dang, J. y Liu, L. (2025). Dehumanization risks associated with artificial intelligence use. American Psychologist. https://doi.org/10.1037/amp0001542

de Mooij, S., Lämsä, J., Lim, L., Aksela, O., Athavale, S., Bistolfi, I., . . . Bannert, M. (2025). A Systematic Review of Self-Regulated Learning through Integration of Multimodal Data and Artificial Intelligence. Educational Psychology Review, 37(2), 54. https://doi.org/10.1007/s10648-025-10028-0

Feng, S., Magana, A. J. y Kao, D. (2021). A systematic review of literature on the effectiveness of intelligent tutoring systems in STEM. 2021 IEEE frontiers in education conference (fie), https://doi.org/10.1109/FIE49875.2021.9637240

Ferhataj, A., Memaj, F., Sahatcija, R., Ora, A. y Koka, E. (2025). Ethical concerns in AI development: analyzing students' perspectives on robotics and society. Journal of Information, Communication Ethics in Society, 23(2), 165-187. https://doi.org/10.1108/JICES-08-2024-0111

Fowler, D. S. (2023). AI in higher education. Journal of Ethics in Higher Education, 3(3), 127-143. https://doi.org/10.26034/fr.jehe.2023.4657

Gao, R. (2025). The Erosion and Reconstruction of Teachers' Authority in the Context of Artificial Intelligence. Journal of Theory Practice in Education Innovation, 2(2), 26-33. https://doi.org/10.5281/zenodo.15117783

George, A. S. (2023). Preparing students for an AI-driven world: Rethinking curriculum and pedagogy in the age of artificial intelligence. Partners Universal Innovative Research Publication, 1(2), 112-136. https://doi.org/10.5281/zenodo.10245675

George, B. y Wooden, O. (2023). Managing the strategic transformation of higher education through artificial intelligence. Administrative Sciences, 13(9), 196. https://doi.org/10.3390/admsci13090196

Gligorea, I., Cioca, M., Oancea, R., Gorski, A.-T., Gorski, H. y Tudorache, P. (2023). Adaptive learning using artificial intelligence in e-learning: A literature review. Education Sciences, 13(12), 1216. https://doi.org/10.3390/educsci13121216

Halkiopoulos, C. y Gkintoni, E. (2024). Leveraging AI in e-learning: Personalized learning and adaptive assessment through cognitive neuropsychology—A systematic analysis. Electronics, 13(18), 3762. https://doi.org/10.3390/electronics13183762

Hutson, J., Jeevanjee, T., Vander Graaf, V., Lively, J., Weber, J., Weir, G., . . . Plate, D. (2022). Artificial intelligence and the disruption of higher education: Strategies for integrations across disciplines. Creative Education, 13(12). https://doi.org/10.4236/ce.2022.1312253

Hyatt, S. E. y Owenz, M. B. (2024). Using universal design for learning and artificial intelligence to support students with disabilities. College Teaching, 1-8. https://doi.org/10.1080/87567555.2024.2313468

Jiang, F., Dong, L., Peng, Y., Wang, K., Yang, K., Pan, C. y You, X. (2024). Large AI model empowered multimodal semantic communications. IEEE Communications Magazine. https://doi.org/10.1109/MCOM.001.2300575

Kotsis, K. T. (2024). Artificial Intelligence Creates Plagiarism or Academic Research? European Journal of Arts, Humanities Social Sciences, 1(6), 169-179. https://doi.org/10.59324/ejahss.2024.1(6).18

78

Artificial Intelligence in Higher Education: Ethical Risks and Transformative Potential. A Systematic Review

Kotsis, K. T. (2025). Artificial intelligence as a catalyst for changes in university-level science education. EIKI Journal of Effective Teaching Methods, 3(3). https://doi.org/10.59652/jetm.v3i3.618

Koukaras, C., Hatzikraniotis, E., Mitsiaki, M., Koukaras, P., Tjortjis, C. y Stavrinides, S. G. (2025). Revolutionising Educational Management with AI and Wireless Networks: A Framework for Smart Resource Allocation and Decision-Making. Applied Sciences, 15(10), 5293. https://doi.org/10.3390/app15105293

Laak, K. J. y Aru, J. (2025). AI and personalized learning. Educational Technology Society, 28(4), 133-150. https://www.jstor.org/stable/48839360

Laflen, A. (2023). Exploring how response technologies shape instructor feedback: A comparison of Canvas Speedgrader, Google Docs, and Turnitin GradeMark. Computers composition, 68, 102777. https://doi.org/10.1016/j.compcom.2023.1 02777

Lin, C.-C., Huang, A. Y. y Lu, O. H. (2023). Artificial intelligence in intelligent tutoring systems toward sustainable education: a systematic review. Smart learning environments, 10(1), 41. https://doi.org/10.1186/s40561-023-00260-y

Neri, H. y Domingos, V. (2023). Navigating the double divide: Generative AI and the dynamics of inequality in latin america. Protosociology: An International Journal of Interdisciplinary Research, 40. https://doi.org/10.5840/protosociology202 34018

Saborío, S. y Rojas, F. J. (2024). Universal design for learning and artificial intelligence in the digital era: Fostering inclusion and autonomous learning. International Journal of Professional Development, Learners Learning, 6(2), ep2408.
https://doi.org/10.30935/ijpdll/14694

Salma, S. S. y Ahmed, S. M. B. (2024). Artificial Intelligence and the Evolution of Educator Roles in the Digital Technology Era. Pakistan JL Analysis Wisdom, 3, 231. https://heinonline.org/HOL/LandingPage?h andle=hein.journals/pknjlolw3&div=239&i d=&page=

Selgas, M. (2025). Ética algorítmica en la educación: un marco integrado para la formación ética de estudiantes mediante sistemas de inteligencia artificial: Algorithmic Ethics in Education: An Integrated Framework for Student Ethics Training Using Artificial Intelligence Systems. INTELETICA. Revista de Inteligencia Artificial, Ética y Sociedad, 2(3), 28-48. https://inteletica.iberamia.org/index.php/jou rnal/article/view/34/14

Siddiqui, M., Hussain, S. A., Saleemi, H. y Fatmi, K. (2025). The Intersection of AI Educational Psychology and Learning Analytics Predicting Student Dropout Risk through Behavioural Indicators. The Critical Review of Social Sciences Studies, 3(3), 104-120. https://doi.org/10.59075/k2zgws84

Singh, B. (2024). INTELLIGENT TUTORING SYSTEMS AND ADAPTIVE LEARNING ENVIRONMENTS: TEACHER-CENTRIC METHOD IN AI-AUGMENTED CLASSROOMS. Asian Education Learning Review, 2(2), 53-68. https://doi.org/10.14456/aelr.2024.10

Singh, B., Kaunert, C., Lal, S. y Arora, M. K. (2025). Enhancing AI-Augmented Classrooms: Teacher-Centric Integration of Intelligent Tutoring Systems and Adaptive Learning Environments. In Fostering Inclusive Education With AI and Emerging Technologies (pp. 99-130). IGI Global. https://doi.org/10.4018/979-8-3693-7255-5.ch004

Skoryk, A., Antipina, I., Havrosh, O., Shunevych, Y. y Shvets, V. (2025). Machine thinking and human imagination: New horizons for creativity in the digital age. International Journal on Culture,

History, Religion, 7(SI1), 115-139. https://doi.org/10.63931/ijchr.v7iSI1.156

Srishti, R. (2024). ChatGPT in Education: Augmenting Learning Experience or Dehumanizing Education? In Educational Perspectives on Digital Technologies in Modeling and Management (pp. 114-128). IGI Global Scientific Publishing. https://doi.org/10.4018/979-8-3693-2314-4.ch005

Strielkowski, W., Grebennikova, V., Lisovskiy, A., Rakhimova, G. y Vasileva, T. (2025). AI-driven adaptive learning for sustainable educational transformation. Sustainable Development, 33(2), 1921-1947. https://doi.org/10.1002/sd.3221

Taylor, C., Dewsbury, B. y Brame, C. (2022). Technology, equity, and inclusion in the virtual education space. In Technologies in Biomedical and Life Sciences Education: Approaches and evidence of efficacy for learning (pp. 35-60). Springer. https://doi.org/10.1007/978-3-030-95633-2_2

Tweedie, M. G. y Sharmi, S. T. (2025). Transparency, assessment and the "Black Box" dilemma. In Navigating Generative AI in Higher Education (pp. 144-159). Edward Elgar Publishing. https://doi.org/10.4337/9781035337873.00013

Vashishth, T. K., Sharma, V., Sharma, K. K., Kumar, B., Panwar, R. y Chaudhary, S. (2024). AI-driven learning analytics for personalized feedback and assessment in higher education. In Using traditional design methods to enhance AI-driven decision making (pp. 206-230). IGI Global Scientific Publishing. https://doi.org/10.4018/979-8-3693-0639-0.ch009

Vasylyshyna, N., Skyrda, T., Lazorenko, N. y Kravets, I. (2024). Legal perspective on Artificial Intelligence andacademic integrity within university educationprocess participants research activity: Newpossibilities along with new limitations. Scientific Journals of the International Academy of Applied Sciences in Lomza, 96(4), 59-82. https://doi.org/10.58246/xr3thd36

Vetrivel, S., Vidhyapriya, P. y Arun, V. (2025). The role of AI in transforming assessment practices in education. In AI Applications and Strategies in Teacher Education (pp. 43-70). IGI Global. https://doi.org/10.4018/979-8-3693-5443-8.ch003

Williamson, B., Macgilchrist, F. y Potter, J. (2023). Re-examining AI, automation and datafication in education. Learning, media technology, 48(1), 1-5. https://doi.org/10.1080/17439884.2023.2167830

Wilson, J., Pollard, B., Aiken, J. M., Caballero, M. D. y Lewandowski, H. (2022). Classification of open-ended responses to a research-based assessment using natural language processing. Physical Review Physics Education Research, 18(1), 010141. https://doi.org/10.1103/PhysRevPhysEducRes.18.010141

Wise, B., Emerson, L., Van Luyn, A., Dyson, B., Bjork, C. y Thomas, S. E. (2024). A scholarly dialogue: writing scholarship, authorship, academic integrity and the challenges of AI. Higher Education Research Development, 43(3), 578-590. https://doi.org/10.1080/07294360.2023.2280195

Wongmahesak, K., Karim, F. y Wongchestha, N. (2025). Artificial Intelligence: A Catalyst for Sustainable Effectiveness in Compulsory Education Management. Asian Education Learning Review, 3(1), 4-4. https://doi.org/10.14456/aelr.2025.4

Yadav, D. (2024). Navigating the landscape of AI integration in education: opportunities, challenges, and ethical considerations for harnessing the potential of artificial intelligence (AI) for teaching and learning. BSSS Journal of Computer, 15(1), 38-48. https://doi.org/10.51767/jc1503

80

Artificial Intelligence in Higher Education: Ethical Risks and Transformative Potential. A Systematic Review

Yeo, M. A. (2023). Academic integrity in the age of Artificial Intelligence (AI) authoring apps. Tesol Journal, 14(3), e716. https://doi.org/10.1002/tesj.716