# Ethical Risks and Governance Frameworks: Exploring the Boundaries of Artificial Intelligence in Educational Applications

**Xiaoming Li**

*Universitas Pendidikan Ganesha, Bali*

*Corresponding Author: *18235656566@139.com*

## ABSTRACT

Artificial Intelligence in educational applications has become a critical issue in global educational development, with its technological potential continuously unfolding. However, this process is accompanied by complex challenges and ethical risks, making systematic ecosystem governance urgently necessary. This study firstly identified the boundaries of AI educational applications, spanning from knowledge acquisition to competency development, and extending to emotional and values cultivation. Secondly, it explored three categories of ethical risks in AI educational applications (including agency risks, safety risks, and development risks) and conducted an attribution analysis focusing on four ecosystem actors (government, AI technology developers, educational institutions, and enterprises). This revealed multi-layered causal chains for the three risk categories, involving four logics: institutional logic, reflecting deficiencies in legal and professional constraints; technical logic, highlighting inherent limitations of algorithms and systems; educational logic, addressing practical challenges of upholding principles versus clinging to tradition; commercial logic, where efficiency and profit distort values. To address these issues, the study proposed an ecosystem governance framework for AI education applications, aiming to provide policy guidance for the symbiotic and sustainable development of AI and education.

**Keywords**: Artificial Intelligence, Educational Application, Ethical Risks, Ecosystem Governance

## INTRODUCTION

The deep integration of Artificial Intelligence (AI) with education has become a significant global trend in educational development. The relationship between AI and education, representing a new wave of technological advancement, has entered a new phase of systematic convergence. AI in educational applications should neither be a one-way “AI empowering education” approach nor a technology-dominated “AI+education” integration model (Harry, 2023). Instead, it should fully leverage the proactive role of educational stakeholders, guided by the genuine needs of educational practice, to steer the development and application of AI technologies. This will achieve a dual-helix co-evolution between education and AI. Such deep integration necessitates a comprehensive governance-level review of AI in education. It is essential to fully consider the unique characteristics of educational settings to meet the demands of deep AI-education integration. To achieve this goal, we must thoroughly analyze the new constraints and risks arising from AI in educational applications, uncover the underlying causes of these risks, and establish an ecosystem governance framework specifically designed to support AI-enabled educational innovation.

## BOUNDARY OF AI EDUCATIONAL APPLICATIONS

The application of Artificial Intelligence (AI) in education refers to the use of technologies such as machine learning, deep learning, and natural language processing to collect, analyze, and provide feedback on educational data. This enables intelligent and personalized support across teaching, learning, assessment, and management processes. In this process, AI not only serves as a tool to assist teaching but is increasingly becoming a generator of learning content, a planner of learning pathways, and a participant in educational decision-making, thereby profoundly influencing the overall operational logic of the education system.

However, education itself is a human-centered value activity whose core objective lies not only in knowledge transmission but also in fostering holistic human development. This characteristic dictates that AI's application in education inherently possesses boundaries: while its functions can replace certain educational processes, it cannot fully supplant the teacher's pedagogical agency, the student's autonomy, or the emotional and value dimensions inherent in educational activities. In other words, AI integration does not signify the automation or dehumanization of education but should be understood as a process of "intelligence augmentation."

From the perspective of educational practice, AI applications in education broadly encompass three levels: Knowledge Transmission Level, Enabling precise delivery of explicit knowledge through intelligent recommendations, knowledge graphs, and online Q&A systems. Competency Development Level, Fostering continuous improvement in students' competencies through personalized learning analytics, behavior tracking, and real-time feedback. Emotional and Value Level, Integrating emotion recognition and motivational functions into teaching via affective computing and human-machine interaction. However, as educational objectives shift from "knowledge acquisition" to "core competencies," AI reveals significant limitations in the third dimension: profound humanistic elements like emotions, empathy, and value judgments remain difficult to translate into genuine educational functions through algorithms and data modeling. Therefore, when exploring the boundaries of AI in education, it is essential to conduct a comprehensive analysis of its applicability and limitations across three dimensions: cognition, emotion, and socialization. (Tuomi, 2022; Palmquist, Sigurdardottir & Myhre, 2025; Chee, Ahn & Lee, 2025) (as illustrated in Figure 1).

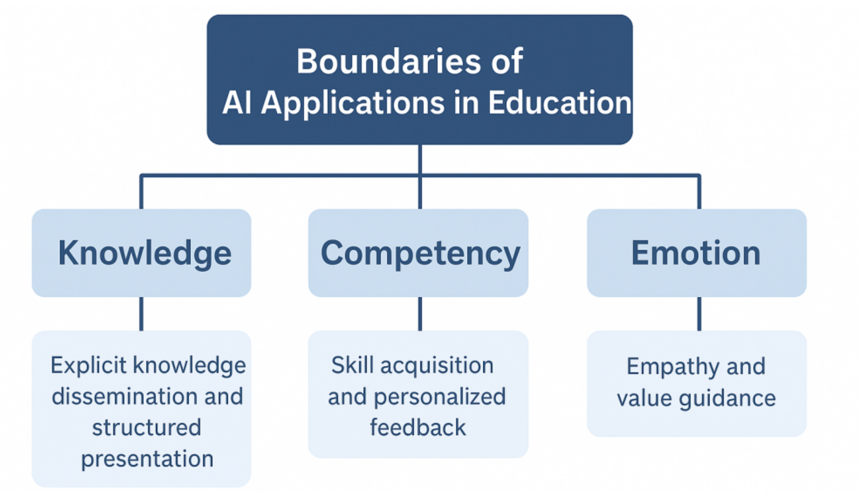


Fig. 1 Boundaries of AI applications in education

**Knowledge level: transmission and structured presentation of explicit knowledge**

In the realm of knowledge acquisition and transmission, artificial intelligence demonstrates distinct technological advantages. Leveraging robust knowledge graphs and natural language processing capabilities, AI achieves efficient organization, updating, and reproduction of explicit knowledge (Jonassen, Yacci & Beissner, 2013; Herschel, Nemati & Steiger, 2001). For instance, intelligent teaching systems can automatically generate personalized learning paths based on learners' progress and cognitive profiles, enabling tailored knowledge delivery. This makes AI significantly superior to traditional teachers in information integration, knowledge retrieval, and logical reasoning. However, this capability remains largely confined to structured knowledge, struggling to encompass the implicit, experiential, and contextual knowledge inherent in educational processes.

**Competency level: skill acquisition and personalized feedback**

In skill development, artificial intelligence enables refined and visualized teaching processes through continuous data analysis and feedback mechanisms. Particularly in learning motor skills and procedural knowledge, AI can decompose instructional tasks into granular steps while providing real-time error correction and multimodal feedback, thereby enhancing learners' operational accuracy and learning efficiency. For instance, AI has demonstrated high instructional effectiveness in medical simulation training, speech pronunciation correction, and artistic technique practice. However, AI remains limited in cultivating complex cognitive skills. It struggles to replace teachers' role in facilitating contextual guidance and intellectual stimulation for developing higher-order abilities like creative thinking and critical judgment.

**Emotional level: limitations in empathy and value guidance**

Education's core lies not only in knowledge transmission but also in emotional development and value cultivation (Martinez, 2014). While AI can simulate learners' emotional states through affective computing and adjust interaction methods, this "emotion" remains an algorithmically simulated representation, lacking genuine emotional experience and moral judgment. Emotional exchange, empathetic resonance, and value guidance in education all involve deep interpersonal understanding and moral perception. AI, lacking consciousness and emotional experience, cannot genuinely assume the role of an "emotional educator." Therefore, in the realm of emotional, attitudinal, and values education, AI can only serve as an auxiliary tool to help teachers identify students' emotional states and learning motivations, not as a primary agent.

**Ethical boundaries in education: human-machine collaboration and redefining the teacher's role**

The boundaries of AI applications in education stem not only from technical capabilities but also from ethical and responsibility concerns. The core value of teachers in the educational process lies not merely in knowledge transmission but also in providing emotional support and guiding values. AI's involvement demands a redefinition of the essence of "teaching" and "nurturing": AI can facilitate knowledge and skill transfer at the 'teaching' level, but human educators must remain central to the "nurturing" dimension. Future education should evolve toward human-AI collaboration (Brusilovsky, 2024), leveraging AI's cognitive and analytical strengths while upholding educational ethics and humanistic values, thereby preserving education's human-centered nature and social functions.

**RISK ASSESSMENT OF AI IN EDUCATIONAL APPLICATIONS**

Against the backdrop of AI in educational applications, artificial intelligence is subtly reshaping traditional teacher-student relationships and learning paradigms. Technology-based power structures may reinforce AI's "authoritative" role in education, thereby amplifying inherent flaws in its "black-box" algorithms and exposing human ethical shortcomings. Ultimately, this poses multifaceted challenges to the educational ecosystem (as illustrated in Figure 2).

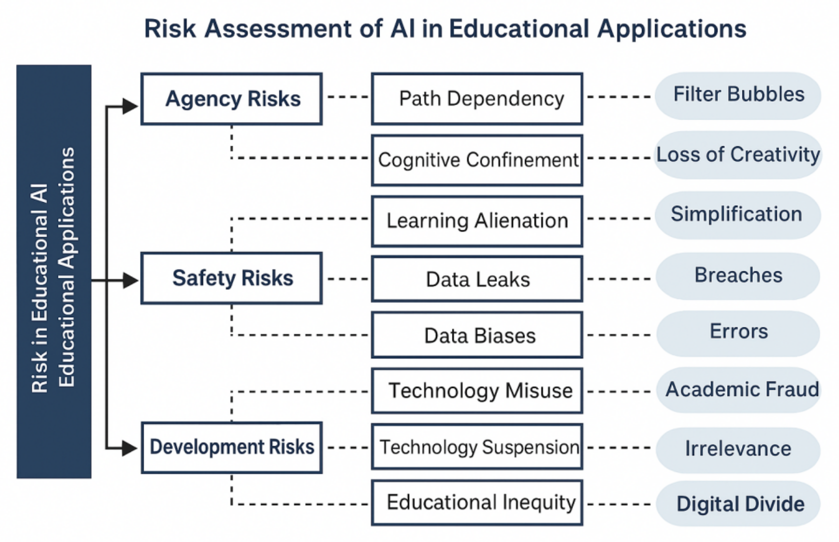


Fig. 2 Risks assessment of AI in educational applications

**Agency Risk**

First, AI, as a powerful and novel "external brain," provides learners with convenient channels for knowledge acquisition, but it also introduces the risk of path dependency. Relying on big data and high computing power, AI's core function lies in generating and pushing relevant learning information through computational pathways and statistical probabilities. This algorithm-driven embedding model superficially enhances learning intelligence and precision, yet subtly reshapes learners' cognitive pathways. Overreliance on AI systems gradually subjects students' learning decisions to AI's computational choices rather than self-directed judgments rooted in intrinsic thinking.

Second, AI standardizes learners' cognitive paths through "centralized" commonality features and "guided" intelligent recommendations. Overuse risks cognitive confinement. AI's "generation" is grounded in group commonality assumptions, fundamentally constructing a unified "center" through algorithms to represent complex human collectives. Students trapped within the information loop governed by algorithmic frameworks receive content aligned with their existing preferences, fostering a path of "pre-established cognition." This triggers the effects of "filter bubbles" and "information silos" within the cognitive field, restricting individual cognitive freedom and creativity (Tomassi, Falegnami & Romano, 2024).

Third, AI algorithmic models dominate the representation of numerous learning scenarios, and their improper use may lead to risks of learning alienation. As O'Neill observes, "No model can capture all the complexities of the real world or all the nuances of human interaction" (O'Neill, 2008). Relying solely on highly simplified algorithmic models struggles to fully restore the essence of education and may even fragment its complexity (Gulson, Sellar & Webb, 2022), manifesting primarily in two dimensions: learning contexts and teaching interactions. Regarding learning contexts, certain AI technologies designed for exam-oriented education may be misused to reinforce standardized testing and mechanized drills, neglecting the humanistic attributes of education (Yang, et al., 2021) and leading to the simplification and formulaic nature of educational activities. In terms of teaching interactions, AI's emotional support relies on pre-set algorithmic logic, making it difficult to genuinely replace the nuanced care and value guidance provided by human teachers in complex emotional exchanges.

**Safety Risks**

First, data leakage risks. The development and application of AI technology require massive amounts of data, introducing potential risks of data breaches. Particularly in educational settings, interactions between teachers, students, and AI involve substantial personal information and learning data. Without robust mechanisms for data encryption, access control, and protection, sensitive data may be improperly accessed or leaked, posing serious security threats to users (Herath, et al., 2024).

Second, data bias risks. AI relies on extensive datasets and developers' design philosophies. If these contain biases related to ethnicity, geography, or other factors, they may generate educational content that deviates from the cognitive frameworks and values of students. Given that database information may contain factual errors or conceptual confusion, AI systems often generate erroneous or nonsensical information. "AI hallucinations" can potentially mislead students (Yingzhe, 2025). Moreover, the content generation process lacks transparency and explainability, functioning like a "black box." Consequently, such misinformation is difficult to identify promptly and correct accurately.

Third, risks of technological misuse. Students can easily exploit AI for academic cheating—such as obtaining assignment answers, generating essay content, or even writing code—thereby bypassing deep knowledge comprehension and independent learning processes (Ganiyu, 2025). Such abuses challenge fundamental principles of academic integrity. Beyond this, concerns exist about students using AI for criminal activities, including defrauding peers or spreading misinformation.

**Development Risks**

First, the risk of technological suspension. AI technology is progressively empowering an increasing number of educational contexts, yet its applicability faces a "suspension dilemma" (Liu, 2025), resulting in a state of "attempting to approach yet failing to integrate"—a detached, suspended condition. Specifically, while it can handle information integration, search, and basic recommendations, and offer rudimentary support for superficial personalized learning, it lacks a deep understanding of educational complexity. It struggles to accurately identify and respond to teachers' dynamic needs in instructional design and classroom management. Furthermore, existing technology falls short in addressing deeper student needs such as learning motivation, metacognitive abilities, and long-term academic development.

Second, risks to educational equity. The widespread adoption of AI technology requires robust digital infrastructure and resource support. Access to these resources varies significantly across regions. In rural and economically underdeveloped areas, weak digital development capabilities prevent students from receiving AI educational support equivalent to that in urban, developed regions, leading to unequal distribution of educational resources. The complexity and high barriers to entry of the technology further widen the gap between different groups. In this process, the "digital divide" becomes increasingly hidden or even rationalized. Marginalized groups struggle to recognize their own disadvantages, forming an invisible group of "digital refugees" (Potocky, 2021).

Third, the risk of excessive surveillance. AI's real-time monitoring of behavior and performance in educational settings places both teachers and students within a highly transparent "data surveillance" environment, subjecting teaching processes to meticulously quantified assessments (Andrejevic & Gates, 2014). For teachers, if every action is tied to evaluation metrics, technological support may transform into technological pressure, thereby diminishing pedagogical creativity and flexibility. For students, continuous AI tracking may induce learning under the pressure of surveillance, turning education into a task subject to scrutiny and quantification rather than an intrinsically driven exploration process. These risks weakening both learning motivation and creative potential.

## ATTRIBUTION ANALYSIS OF RISKS IN AI IN EDUCATIONAL APPLICATIONS

The emergence of risks in AI in educational applications is not driven by a single factor but is deeply rooted in the absence of responsibility and interaction among diverse stakeholders. Within the ecosystem of AI in educational applications, governments, AI technology developers, schools, and enterprises constitute four core entities. The underlying institutional logic, technological logic, educational logic, and commercial logic embodied by each respectively harbor potential risks. These four interwoven logics collectively form a multi-layered causal chain of risks in AI in educational applications (as illustrated in Figure 3).

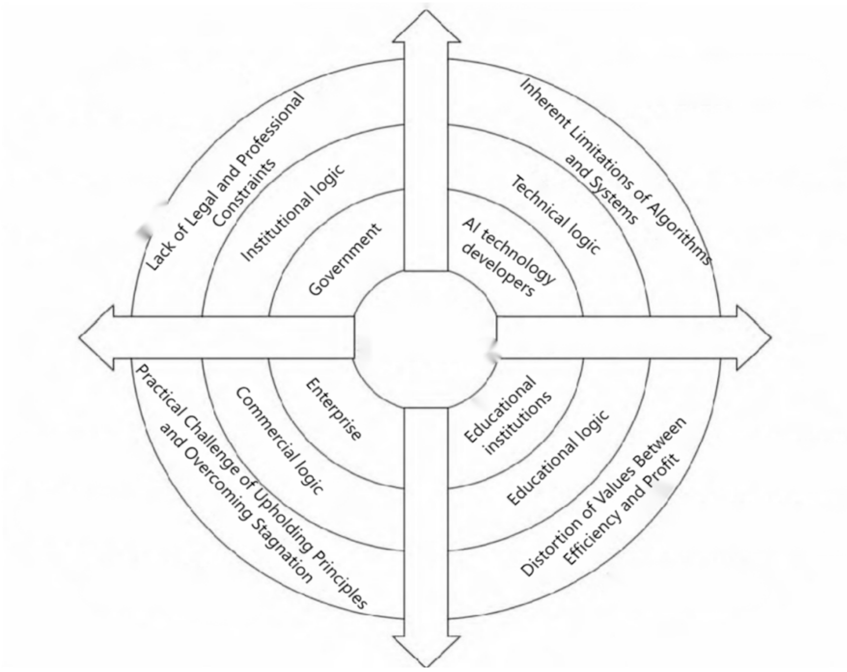


Fig. 3 Attribution of risks in AI educational applications

### Institutional Logic: Lack of Legal and Professional Constraints

China's data governance framework primarily rests on the Data Security Law of the People's Republic of China, the Cybersecurity Law of the People's Republic of China, and the Personal Information Protection Law of the People's Republic of China. However, significant gaps persist regarding the specific application of AI in educational settings, which fundamentally contributes to the risks associated with AI in educational applications.

At the legal level, the ambiguity in the division of responsibilities within the current governance framework and policies has become a significant source of risk. Existing policies, such as the Interim Measures for the Administration of Generative Artificial Intelligence Services, primarily focus on ensuring the compliant operations of technology companies. However, they fail to effectively allocate responsibilities among key stakeholders in educational settings—including schools, technology companies, and government agencies—resulting in unclear regulatory and enforcement accountability. Furthermore, the existing policy framework has yet to establish a tiered, multi-stakeholder collaborative governance system. Taking educational data oversight as an example, while some regions have experimented with involving both enterprises and schools in regulatory efforts, the absence of clearly defined hierarchical oversight responsibilities often renders such collaborations superficial, failing to form an effective closed-loop risk governance system (Cameron et al., 2011).

At the professional level, specialized governance institutions and institutional designs tailored to educational contexts are notably absent. The existing legal framework primarily follows a general-purpose data and technology governance orientation, struggling to fully meet the educational sector's requirements for student development, diversity, and equity. This generalized governance framework also fails to effectively cover the specific details involved in AI educational applications. The unique nature of students as the primary subjects necessitates special regulations when AI is applied to student populations. For instance, personalized learning for students cannot adopt the underlying logic of consumer-oriented personalized recommendations. It necessitates clear standards and norms grounded in educational expertise and extensive scientific empirical research. Furthermore, to address the current management dilemma of "over-regulation stifling innovation while lax oversight breeds chaos," governments must swiftly mobilize professionals to conduct scientific research on new products and applications, establishing diversified standards.

**Technical Logic: Inherent Limitations of Algorithms and Systems**

The current immaturity of AI technology has given rise to inherent instability and complexity issues. These challenges permeate multiple stages of AI system development—including research, design, manufacturing, and application—and constitute the intrinsic root causes of risks associated with AI in educational applications.

Unstable factors such as algorithmic black boxes and data bias pose significant challenges to the safety and controllability of AI applications. Unlike traditional rule-based algorithms, modern machine learning algorithms possess self-learning and autonomous decision-making capabilities, enabling them to extract patterns from massive datasets without direct human intervention. However, this data-driven decision-making process, due to its "black box" nature, makes it difficult for outsiders to understand the logic from input to output, complicating safety verification (Corso et al., 2021). Data serves not only as the critical foundation for AI optimization but also as the key to achieving fair and reasonable decision-making (Chen, Wu & Wang, 2023). Poor data quality inevitably degrades the performance and reliability of AI systems.

Ethical definitions surrounding AI technology reveal deep theoretical disagreements, complicating the confirmation of responsibilities and authority in governance. Traditional perspectives adhere to a subject-object dualism, viewing AI as an auxiliary tool for achieving educational objectives and emphasizing the immutable dominance of human agency (Mouta et al., 2025). Another perspective advocates for intersubjective relationships, positing that AI possesses potential agency within intelligent and automated environments, potentially eroding humanity's exclusive status as the primary educational agent (Papa & Jackson, 2021). With no consensus reached between these opposing views, AI's ethical standing remains undefined. This ambiguity may hinder precise attribution of AI responsibility within legal and ethical frameworks, potentially leading to deeper governance issues such as responsibility shifting and ethical vacuums.

The rapid advancement of AI technology and its expanding applications in education further intensify governance challenges and risks. Currently, AI applications in education—ranging from personalized recommendation systems to intelligent teaching assistants—are permeating diverse scenarios including curriculum design, instructional feedback, and student performance analysis, with their functions and influence continually expanding. However, the rapid iteration of technology often outpaces regulatory mechanisms, resulting in a lack of unified standards for AI system implementation in education and increasing susceptibility to uncontrolled risks (Sanyal et al., 2024).

**Educational Logic: The Practical Challenge of Upholding Principles and Overcoming Stagnation**

Currently, educational reform has entered a critical phase. The complex interplay between its pursuit of upholding principles while fostering innovation and the persistent structural challenges it faces collectively form the root causes constraining AI in educational applications at the governance level.

School education emphasizes upholding principles while pursuing innovation—that is, maintaining educational values and social responsibilities while aiming to cultivate moral character and nurture innovative talent through holistic innovation integrating moral, intellectual, physical, aesthetic, and labor education. This requires a cautious approach to adopting new technologies (Khanagha et al., 2013), evaluating their legitimacy and appropriateness from the perspective of student development to prevent one-dimensional technological transformation of education. However, despite numerous innovative applications and models emerging in AI's integration into education, empirical research on the safety and effectiveness of these technologies in teaching remains limited. This lack of comprehensive evidence undermines the justification for their long-term application and safety, while also making it difficult to ensure these applications align with core educational values and societal expectations. Consequently, schools face significant trial-and-error and adjustment costs in AI implementation.

The persistent tendency toward exam-oriented education within the current system poses systemic constraints on the application and governance of AI in educational applications. Certain AI educational products, such as smart homework platforms and online assessment systems, primarily focus on improving exam scores as their core objective. By pinpointing students' weaknesses and providing repetitive drills, they optimize academic performance in the short term. This “drill-and-kill” approach reduces the learning process to mere test preparation, neglecting the cultivation of critical thinking, inquiry skills, and interdisciplinary abilities. Simultaneously, AI systems’ effectiveness evaluations overly rely on quantitative metrics like accuracy rates and test scores. This aligns with education’s excessive focus on academic achievements, further driving teachers and schools to prioritize grades over holistic student development. This model makes “cutting-edge technology” serve “backward education,” not only contradicting the goal of integrating moral, intellectual, physical, aesthetic, and labor education but also hindering AI’s long-term development in education.

The lagging development of digital literacy among teachers and students represents another pressing issue in current educational advancement. As the core participants in education, the relatively underdeveloped digital literacy and technological adaptability of educators and learners in a rapidly evolving technological environment not only slows the pace of deep integration between technology and education but also exposes deficiencies in the education system's capacity to cultivate these competencies.

**Commercial Logic: The Distortion of Values Between Efficiency and Profit**

There exists an inherent tension between the commercial nature of enterprises as market entities and the public welfare nature of education (Molnar, 2006). AI education companies often prioritize commercial interests, neglecting respect for educational principles and the realization of long-term educational goals. This represents an exogenous risk factor in AI in educational applications. Within market-driven environments, education enterprises typically prioritize maximizing commercial profits in product development and service delivery. This profit-seeking motivation leads companies to design products that cater to market demands, pursuing rapid growth in user numbers and revenue while neglecting educational principles and objectives. The widespread lack of solid educational theory foundations among AI education enterprises further exacerbates governance risks. Many companies' R&D teams are primarily composed of personnel from computer science, data science, and similar fields, with relatively weak understanding of disciplines like pedagogy and psychology.

This technology-driven development model easily leads to neglect of education's essence in product design, making AI education products fundamentally ill-suited to meet educational needs. Fierce market competition within the industry often intensifies short-sighted behavior among companies, creating fertile ground for risks. Some companies pursue "policy-driven innovation," focusing on superficial compliance while neglecting deep-level design innovation (Bamber, 2004). To rapidly capture market share, many adopt a "launch first, optimize later" strategy, prioritizing speed over continuous refinement of product quality and user experience. This approach exposes numerous issues during actual use, such as algorithmic instability and inadequate data privacy protection.

## AN ECOSYSTEM GOVERNANCE FRAMEWORK FOR AI EDUCATION APPLICATION

Based on risk assessment and attribution analysis, this study proposes an AI-assisted ecosystem governance framework for educational applications to address the opportunities and challenges of deep integration. The ecological governance of AI-assisted educational applications should center on upholding educational values, adhering to a governance approach where the government serves as the core, with collaborative participation from enterprises and educational users. Through tripartite coordination, an information circulation and feedback mechanism is established to ensure the governance system continuously optimizes according to dynamic demands, achieving autonomous evolution and dynamic equilibrium within the ecosystem. This governance framework emphasizes three key dimensions: refining the top-level architecture, standardizing industry management, and enhancing stakeholder adaptability.

**Enhance government-led scientific management and collaborative governance**

As a public good, the AI-driven transformation of education cannot rely solely on market forces. It requires comprehensive government guidance through macroeconomic regulation and policy formulation. Specifically, the government should provide systematic solutions for risk governance in AI-enabled educational transformation by focusing on three dimensions: establishing regulations, implementing collaborative governance, and conducting pilot experiments.

First, scientifically establish a legal foundation to mitigate risks associated with AI in educational applications. On one hand, accelerate the refinement of privacy protection and risk prevention regulations concerning educational data. Clearly define lifecycle management strategies for educational data, standardize intellectual property rights attribution (Hollmann et al., 2022), and establish periodic review mechanisms to ensure lawful data usage and controllable risks. Second, establish a market access review system encompassing industry standards and product specifications. This system should set clear requirements for the educational value, algorithmic transparency, and data security of AI education products, ensuring technology providers have clear guidelines during product development and service delivery. Consider establishing an access review body for AI products in education, drawing on the experimental and review processes for new drug approvals to ensure effective risk prevention for AI products in the education sector.

Second, multi-stakeholder collaborative governance serves as the key pathway for government to achieve top-level guidance. The government should leverage its coordinating role to promote deep engagement of multiple entities in the education sector within governance frameworks, forming a dynamic, interconnected risk prevention mechanism. To this end, a specialized technical regulatory department for AI applications in education should be established. This department must possess professional regulatory theories, rules, methodologies, technologies, and processes to conduct oversight in accordance with the law, radiate influence internationally, and provide public services for "AI-powered educational transformation." Positioned as both a regulator safeguarding educational security and equity, and a key player in international cooperation and national governance, this department should spearhead the construction of an intelligent education data supervision system. It should integrate the capabilities of government, schools, and relevant enterprises to advance collaborative oversight of educational data risks.

Third, pilot experiments represent a crucial practical strategy for advancing AI in educational applications under top-level government guidance. Authorities should adopt an evidence-based "pilot-to-scale" exploration model, thoroughly validating and evaluating small-scale pilot projects before gradual expansion. This approach provides scientific grounds for educational reform through empirical research and data support, mitigating potential risks from hasty implementation while ensuring the feasibility, significance, and effectiveness of new technology applications.

**Standardize industry-wide technical and qualification management alongside industry self-regulation**

As highlighted by the "Collinridge Dilemma," when the societal consequences of a technology remain unclear in its early stages, failure to implement effective preventive measures can lead to significant challenges in governance once adverse outcomes become entrenched within economic and social structures (Tierney, 2012). Therefore, it is crucial to implement stringent safeguards against potential risks before AI technology inflicts harm on the education sector. To standardize industry management, systematic development should advance through three dimensions: comprehensive technical specifications across the entire industrial chain, professional certification examinations, and industry associations.

First, establishing comprehensive technical specifications across the entire industrial chain is a key measure for achieving corporate governance in AI in educational applications. The National Artificial Intelligence Industry Comprehensive Standardization System Construction Guide (2024 Edition) divides the AI industry chain into four segments: the foundation layer, framework layer, model layer, and application layer, covering all stages from underlying architecture to practical implementation. Considering the entire process of educational product development, industry management standards can be established across these four tiers. At the foundation layer, algorithmic transparency and explainability should be enhanced, alongside implementing rigorous data governance frameworks to safeguard user privacy and data security. At the framework layer, security testing for open-source frameworks must be strengthened, compatibility standards established, and secure, efficient technical development environments built. At the model layer, efforts should focus on eliminating algorithmic bias and ethical risks, while promoting the research, development, and application of localized large models to ensure dual safeguards of fairness and adaptability. At the application layer, a dynamic risk assessment system should be established, and resource allocation mechanisms should be refined to ensure the stability and inclusivity of AI technology in educational settings, thereby advancing educational equity and intelligent development.

Second, establish a professional certification and qualification examination system for AI in educational applications. As a cross-disciplinary field integrating artificial intelligence and education, AI education demands higher professional standards from practitioners. They must not only possess core AI technology development capabilities but also systematically master fundamental theories in education, psychology, and related fields. Through certification and qualification examinations, clear competency standards can be established to ensure practitioners possess both theoretical grounding and practical capabilities in technology development and educational application. Furthermore, requiring edtech companies to establish teaching and research departments and making professional certification a prerequisite for teaching R&D team membership is a crucial method to enhance the scientific rigor of industry governance. This approach enables rigorous evaluation of AI technology's suitability for educational contexts, mitigating risks arising from misuse or bias.

Third, establishing an industry association for AI in educational applications is a critical pathway to promote industry standardization and high-quality development. As a vital platform connecting enterprises, educational institutions, and regulatory bodies, such associations play an irreplaceable role in standardization, oversight, and resource integration. Firstly, the association can spearhead the development of industry standards covering algorithm transparency, data ethics norms, and educational scenario adaptability. This provides clear guidance for corporate R&D and technology application, ensuring deep integration between AI technology and educational needs. Second, associations should advance professional certification and training systems to elevate practitioners' expertise and overall industry standards. Additionally, they can regularly convene discussions on core AI education issues to build industry consensus and provide scientific support for policy formulation. Simultaneously, by strengthening industry self-governance, associations can promote corporate self-regulation mechanisms and establish dynamic monitoring and feedback systems to promptly address risks and challenges in technology deployment.

**Leading role of educational entities in values, competencies, and innovation**

AI-powered human-machine collaborative learning in educational transformation relies not only on system optimization but also on the proactive engagement of educational stakeholders. To fulfill education's guiding role, these stakeholders must provide robust support at the institutional level through innovations in values, competencies, and methodologies.

First, uphold educational values as the guiding principle for technological innovation. "Human-centered, AI for good" has become a shared principle in global technology governance. This principle is not only a fundamental requirement of technological ethics but also the foundational value ensuring AI serves the essence of education. As O'Neill states, algorithms should not be endowed with omnipotence. Any technological utopian fantasy that views them as the ultimate solution to education's complexity and diversity deviates from the core purpose of education. Outstanding education scholars and frontline practitioners must uphold educational values, applying AI to foster holistic human development rather than succumbing to technocratic fervor. Educational institutions like schools should strengthen value guidance through institutionalized measures—such as specialized training, workshops, or teaching case analyses—to enhance faculty and student awareness of technological ethics. This cultivates a mutually beneficial relationship between technology and education.

Second, enhance AI literacy to ensure mastery of technology. The AI literacy of educational users (teachers, students, parents) serves not only as a crucial barrier against technological risks but also as the core force ensuring technology's effective application in education. AI literacy concerns the adaptability, judgment, and agency of educational users within the technological ecosystem. Its scope extends beyond mastering basic technical functions to encompass critical attitudes, ethical awareness, and practical skills in AI technology use [30]. Specifically, educators should deeply understand technology's functions and limitations in education, effectively adapt its application in teaching practices, and proactively identify and report technological risks. Students should possess foundational knowledge and application skills regarding AI technology, ensuring they maintain critical thinking and ethical awareness during use. Parents should develop the ability to select and supervise AI educational products, ensuring minors receive appropriate protection and guidance while using technology.

Third, provide essential support for teachers' proactive innovation. As AI technology deeply integrates into education, actively exploring new methods and pathways is crucial for mitigating technological risks and fostering educational innovation. Teachers should boldly experiment in their teaching practices, integrating AI technology into diverse scenarios such as classroom instruction, assignment design, and student performance analysis. They should explore the challenges and limitations of AI in educational applications and provide improvement suggestions to school administrators and technology vendors through feedback mechanisms. Simultaneously, teachers can establish professional learning communities to share exploration outcomes and practical experiences, fostering collective intelligence that drives technological innovation. Additionally, schools and educational institutions should encourage students to maintain critical thinking when using AI tools and actively participate in technology evaluation and feedback processes. Regularly organizing activities such as student opinion surveys, experience feedback sessions, or innovation competitions can provide multidimensional user perspectives for technological refinement.

## CONCLUSION

The rapid advancement of artificial intelligence is embedding itself into education systems at an unprecedented pace. While injecting new momentum into educational reform, it also challenges the foundational values and governance models of traditional education. From the "instrumental rationality" of technological integration to the "ecological governance" of systemic reconstruction, the future landscape of AI in educational applications is no longer merely a matter of technical deployment. Instead, it requires concerted efforts from multiple stakeholders—guided by educational principles, proactive engagement, and the formation of new collaborative mechanisms and governance frameworks.

The education of the future should not be defined by algorithms, but rather by education defining the trajectory of technological advancement. Advancing the systemic ecological governance of AI in educational applications represents not only a profound response to the essence of education in the intelligent era, but also a crucial step toward achieving equitable, high-quality education in the years ahead.

## REFERENCES

Andrejevic, M., & Gates, K. (2014). Big data surveillance: Introduction. Surveillance & Society, 12(2), 185-196.

Bamber, V. (2004). An Exploration of Policy Enactment, through the study of the introduction of lecturer development programmes. Lancaster University (United Kingdom).

Brusilovsky, P. (2024). AI in education, learner control, and human-AI collaboration. International Journal of Artificial Intelligence in Education, 34(1), 122-135.

Cameron, F., Bequette, B. W., Wilson, D. M., Buckingham, B. A., Lee, H., & Niemeyer, G. (2011). A closed-loop artificial pancreas based on risk management. Journal of diabetes science and technology, 5(2), 368-379.

Chee, H., Ahn, S., & Lee, J. (2025). A competency framework for AI literacy: Variations by different learner groups and an implied learning pathway. British Journal of Educational Technology, 56(5), 2146-2182.

Chen, P., Wu, L., & Wang, L. (2023). AI fairness in data management and analytics: A review on challenges, methodologies and applications. Applied sciences, 13(18), 10258.

Corso, A., Moss, R., Koren, M., Lee, R., & Kochenderfer, M. (2021). A survey of algorithms for black-box safety validation of cyber-physical systems. Journal of Artificial Intelligence Research, 72, 377-428.

Ganiyu, T. O. (2025). Academic integrity in the AI era: Battling cheating with innovation. AI and Ethics, Academic Integrity and the Future of Quality Assurance in Higher Education.

Gulson, K. N., Sellar, S., & Webb, P. T. (2022). Algorithms of education: How datafication and artificial intelligence shape policy. U of Minnesota Press.

Harry, A. (2023). Role of AI in education. Interdiciplinary Journal & Hummanity *(INJURITY*), 2(3).

Herath, H. M. S. S., Herath, H. M. K. K. M. B., Madhusanka, B. G. D. A., & Guruge, L. G. P. K. (2024). Data protection challenges in the processing of sensitive data. In Data Protection: The Wake of AI and Machine Learning (pp. 155-179). Cham: Springer Nature Switzerland.

Herschel, R. T., Nemati, H., & Steiger, D. (2001). Tacit to explicit knowledge conversion: knowledge exchange protocols. Journal of knowledge management, 5(1), 107-116.

Hollmann, S., Regierer, B., D'Elia, D., Kisslinger, A., & Liguori, G. L. (2022). Toward the definition of common strategies for improving reproducibility, standardization, management, and overall impact of academic research. In Advances in Biomembranes and Lipid Self-Assembly (Vol. 35, pp. 1-24). Academic Press.

Jonassen, D. H., Yacci, M., & Beissner, K. (2013). Structural knowledge: Techniques for representing, conveying, and acquiring structural knowledge. Routledge.

Khanagha, S., Volberda, H., Sidhu, J., & Oshri, I. (2013). Management innovation and adoption of emerging technologies: The case of cloud computing. European Management Review, 10(1).

Liu, L. (2025, May). Advancing Teachers' Digital Competence in Western China's Inclusive Kindergartens Through Human-Machine Collaboration: An Intelligent Diagnosis-Driven Pathway for Integrating Local Resources. In 2025 5th International Conference on Artificial Intelligence and Education (ICAIE) (pp. 299-303). IEEE.

Martinez, M. E. (2014). Education as the cultivation of intelligence. Routledge.

Molnar, A. (2006). The commercial transformation of public education. Journal of Education Policy, 21(5), 621-640.

Mouta, A., Pinto-Llorente, A. M., & Torrecilla-Sánchez, E. M. (2025). "Where is Agency Moving to?": Exploring the Interplay between AI Technologies in Education and Human Agency. Digital Society, 4(2), 49.

O'Neill, S. (2008). Approaching Interaction. Interactive Media: The Semiotics of Embodied Interaction, 1-21.

Palmquist, A., Sigurdardottir, H. D. I., & Myhre, H. (2025). Exploring interfaces and implications for integrating social-emotional competencies into AI literacy for education: a narrative review. Journal of Computers in Education, 1-37.

Papa, R., & Jackson, K. M. (Eds.). (2021). Artificial intelligence, human agency and the educational leader. Springer Nature.

Potocky, M. (2021). The role of digital skills in refugee integration: a state-of-the-art review. The International Journal of Information, Diversity, & Inclusion (IJIDI).

Sanyal, S., Sharma, P., & Dudani, C. (2024). A complex adaptive system framework to regulate artificial intelligence (No. 26). Working paper.

Tierney, K. (2012). Disaster governance: Social, political, and economic dimensions. Annual review of Environment and Resources, 37(1), 341-363.

Tomassi, A., Falegnami, A., & Romano, E. (2024). Mapping automatic social media information disorder. The role of bots and AI in spreading misleading information in society. Plos one, 19(5), e0303183.

Tuomi, I. (2022). Artificial intelligence, 21st century competences, and socio-emotional learning in education: More than high-risk?. European Journal of Education, 57(4), 601-619.

Yang, S. J., Ogata, H., Matsui, T., & Chen, N. S. (2021). Human-centered artificial intelligence in education: Seeing the invisible through the visible. Computers and Education: Artificial Intelligence, 2, 100008.

Yingzhe, L. I. (2025). Addressing “Hallucinations” in AI-Generated Content: Strategies for Developing Student Fact-Checking and Information Evaluation Skills. Artificial Intelligence Education Studies, 1(2), 48-62.