International Journal of Educational Sciences and Arts (IJESA)
المجلة الدولية للعلوم التربوية والآداب
Vol. (5), No. (6) June 2026 الإصدار (5)، العدد (6)

Ethical Dimensions and Issues of Artificial Intelligence Applications in Education: A PRISMA-Informed Systematic Literature Review

Ahmed Mohammed Muslihi
Master's student, Computer Science, College of Computer Science, King Khalid University, Kingdom of Saudi Arabia
amuslihi@gmail.com

Abdulmohsen Algarni
Professor, College of Computer Science, King Khalid University, Kingdom of Saudi Arabia

Abstract

Background. Artificial intelligence (AI) is increasingly embedded across K–12 and higher education for assessment, tutoring, feedback, predictive analytics, student support, content generation, and institutional decision-making. The rapid diffusion of generative AI has compressed adoption timelines and intensified pre-existing ethical debates surrounding privacy, fairness, transparency, accountability, learner agency, and academic integrity.

Objective. This PRISMA-informed systematic literature review synthesizes recent peer-reviewed scholarship and policy-oriented evidence concerning the ethical dimensions of AI applications in education, with particular emphasis on the 2023–2025 period.

Methods. The review was conducted in accordance with the PRISMA 2020 reporting framework. Predefined review questions, eligibility criteria, a reproducible search-string template, a screening protocol, and a thematic synthesis approach were applied. Searches drew on Scopus, Web of Science, ERIC, IEEE Xplore, ACM Digital Library, ScienceDirect, and SpringerLink, supplemented by policy literature from UNESCO, the OECD, and the European Union. A thematic synthesis was conducted across recurring ethical domains, and a quality-appraisal plan suited to the heterogeneity of the corpus was specified.

Results. Six interdependent ethical domains were consistently identified across the corpus: (1) privacy and surveillance; (2) fairness, bias, and inclusion; (3) transparency and explainability; (4) accountability and governance; (5) learner and educator agency; and (6) academic integrity. Five cross-cutting tensions characterize the field: personalization versus surveillance, efficiency versus pedagogical autonomy, innovation versus fairness, automation versus accountability, and assistance versus integrity. Persistent gaps include limited longitudinal and comparative empirical work, underrepresentation of K–12 and Global South contexts, insufficient attention to procurement and platform governance, and a tendency to treat heterogeneous AI technologies as interchangeable.

Conclusions. Ethical AI in education should be conceptualized not only as a technical design problem but also as a pedagogical, institutional, and sociopolitical challenge. Responsible adoption requires participatory governance, transparent procurement, meaningful human oversight, redesigned assessment, and context-sensitive policy. Future research should prioritize longitudinal and implementation-focused studies, technology-specific risk–benefit assessment, and stronger representation of underrepresented educational contexts.

Keywords: Artificial Intelligence in Education, Educational Ethics, Generative AI, Data Privacy, Algorithmic Bias, Academic Integrity, AI Governance, Learner Agency, PRISMA.

Highlights

- AI ethics in education must be understood through educational purposes and contexts, not solely through general AI principles or technical performance benchmarks.
- Privacy, fairness, explainability, accountability, autonomy, and integrity operate as interdependent ethical domains rather than as isolated concerns.
- The current evidence base is constrained by a shortage of longitudinal, comparative, K–12, Global South, and implementation-focused studies.
- Responsible adoption requires participatory governance, transparent procurement, meaningful human oversight, assessment redesign, and sustained educator development.

1. Introduction

Artificial intelligence (AI) has emerged as one of the most consequential technological developments shaping contemporary education. AI-enabled systems are now routinely deployed to personalize instruction, automate feedback, identify at-risk learners, support academic advising, translate educational materials, generate lesson plans, evaluate written work, and provide conversational tutoring (Taşkın, 2025). The widespread availability of generative AI tools has further accelerated this transformation by moving AI from specialized technical infrastructures into routine educational practice. Consequently, ethical concerns once confined largely to the artificial intelligence in education (AIED) research community have become central issues in institutional policy, classroom design, professional development, and broader public discourse (Schönberger, 2024).

Ethical considerations are particularly salient in education because schools and universities serve vulnerable populations, shape life trajectories, and carry both developmental and evaluative responsibilities (Eden et al., 2024). When AI systems assess student capabilities, recommend interventions, or mediate access to feedback, they reallocate opportunities and responsibilities in ways that are often opaque (Gore & Dove, 2024). Trade-offs deemed acceptable in commercial or consumer contexts may be inappropriate in educational settings, where learners are not merely users but subjects of formative evaluation. Ethical analysis in education must therefore extend beyond abstract principles of AI governance to engage directly with pedagogy, assessment, inclusion, institutional

authority, and the underlying purposes of education (Mouta, Pinto-Llorente, & Torrecilla-Sánchez, 2024).

Recent reviews highlight both the rapid growth of the field and its persistent unevenness. Nikolic et al. (2024) characterize the higher-education literature on AI adoption as expanding rapidly but lacking robust ethical, methodological, and contextual foundations. Fu and Weng (2024) demonstrate that fairness, privacy, agency, transparency, and non-maleficence are consistently identified as core elements of responsible, human-centered AI in education. Mouta et al. (2024), together with Yan, Liu, and Chau (2025), observe that themes such as privacy, bias, explainability, and inclusion recur across the literature but are frequently treated in a fragmented or inadequately theorized manner. More recent analyses by Wieczorek, Hosseini, and Gordijn (2025) indicate that research is beginning to distinguish among educational levels and stakeholder groups, but still lacks adequate empirical grounding and global representation.

This review synthesizes the ethical dimensions and challenges associated with AI applications in education by critically examining recent peer-reviewed scholarship, comparing major ethical themes, and identifying areas where the evidence base remains weak. Rather than treating ethics as a procedural checklist, the review investigates how ethical concerns arise from specific educational uses of AI and considers the implications of these concerns for future research, policy, and institutional practice.

2. Methods

This review was conducted as a PRISMA-informed systematic literature review (Page et al., 2021), synthesizing recent systematic reviews, meta-reviews, selected empirical studies, and major policy frameworks relevant to the ethics of AI in education. The methodological design followed the PRISMA 2020 reporting framework with respect to the formulation of review questions, eligibility criteria, search strategy, screening procedures, data extraction, and thematic synthesis. Where specific procedural details require completion by the corresponding author prior to submission (for example, exact database export counts), these are explicitly flagged in Appendix A.

2.1 Review questions:

The review was structured around three guiding questions:

- RQ1. Which ethical issues are most consistently identified in recent literature on AI applications in education?
- RQ2. How do these issues vary across educational contexts, stakeholder groups, and types of AI applications?
- RQ3. What gaps remain in the evidence base and in institutional governance of educational AI?

2.2 Eligibility criteria:

Studies were eligible for inclusion if they addressed AI applications in educational contexts and engaged with at least one ethical, governance, pedagogical, or stakeholder-related dimension. Eligible

records comprised peer-reviewed reviews, meta-reviews, selected empirical studies, and major policy frameworks. Records were excluded if they addressed AI ethics outside educational contexts, lacked an identifiable ethical dimension, were purely technical without educational implications, or constituted vendor or promotional materials with no scholarly or policy value. Detailed inclusion and exclusion criteria are summarized in Table (1).

Table (1). Eligibility criteria for inclusion of sources

Criterion	Included	Excluded
Topic	AI applications in education, including adaptive learning, learning analytics, automated assessment, generative AI, AI tutoring, institutional analytics, and decision-support systems.	General AI ethics with no clear connection to educational practice or policy.
Document type	Peer-reviewed review articles, meta-reviews, selected empirical studies, and major international policy or governance documents.	Opinion pieces, editorials, and vendor materials, except where used for limited contextual illustration.
Time period	Primary emphasis on 2023–2025 literature, with foundational works retained where required for conceptual framing.	Older sources without continuing conceptual or policy relevance to the ethical themes under review.
Educational level	K–12, higher education, teacher professional development, and institutional governance contexts.	Non-educational training contexts unless directly relevant to educational AI ethics.
Language and scope	English-language scholarly literature and internationally accessible policy frameworks.	Sources not accessible for verification or falling outside the stated scope.

2.3 Search strategy:

The search strategy was designed to capture peer-reviewed scholarship and policy literature addressing the ethical dimensions of AI in educational settings. Searches were conducted across Scopus, Web of Science, ERIC, Education Source, IEEE Xplore, ACM Digital Library, ScienceDirect, and SpringerLink, supplemented by targeted retrieval of policy documents from UNESCO, the OECD, and the European Union, and by backward and forward citation chasing of key reviews. Google Scholar was used for supplementary citation chasing only.

The core reproducible search string was structured as: *("artificial intelligence" OR "AI" OR "generative AI" OR "large language model" OR "learning analytics" OR "adaptive learning") AND (education OR school OR university OR "higher education" OR "K–12") AND (ethics OR privacy OR fairness OR bias OR transparency OR explainability OR accountability OR autonomy OR "academic integrity" OR inclusion OR governance)*. Search terms were adapted to the syntax of each database. Exact database queries, retrieval dates, and identified record counts are reported in Appendix A and should be confirmed against the corresponding author's search log prior to final submission.

2.4 Screening and selection:

Records were imported into a reference manager, deduplicated, and screened in two stages: (i) title and abstract screening against the eligibility criteria, followed by (ii) full-text screening of records retained at the first stage. Records were excluded for documented reasons, including the absence of an educational focus, the absence of an ethical dimension, duplication, non-scholarly status, or unavailability of full text. The screening flow is reported in summary form in Appendix A and is intended to populate a PRISMA 2020 flow diagram in the published version.

2.5 Data extraction and synthesis:

Given the heterogeneity of the corpus in study design, educational level, technology type, and outcome measures, sources were synthesized thematically rather than statistically. Extracted variables included educational context, AI application, stakeholder group, ethical issue, proposed mitigation, governance implications, and reported evidence gaps. The thematic synthesis prioritized convergent themes across reviews and policy frameworks while critically identifying where claims remain conceptual, aspirational, or weakly supported by empirical evidence. The ethical analysis framework that guided synthesis is summarized in Table (2).

Table (2). Ethical analysis framework used in the thematic synthesis

Ethical domain	Core question	Typical educational AI applications	Governance implication
Privacy and surveillance	What learner data are collected, inferred, stored, shared, or monitored?	Learning analytics, proctoring, adaptive systems, AI assistants.	Data minimization, consent, retention limits, vendor controls, privacy impact assessment.
Fairness, bias, and inclusion	Whose language, culture, disability status, infrastructure, or learning profile is represented?	Predictive analytics, automated assessment, recommender systems, tutoring, translation.	Bias audit, accessibility review, participatory design, contextual validation.
Transparency and explainability	Can students, educators, and administrators understand and contest AI outputs?	Dashboards, feedback tools, risk scores, grading support, chatbots.	Disclosure, interpretable outputs, appeal routes, technical and pedagogical documentation.
Accountability and governance	Who is responsible when AI causes educational harm or allocates opportunity inequitably?	Procurement, institutional analytics, advising, grading, disciplinary processes.	Role allocation, audits, human oversight, incident reporting, dispute procedures.
Agency and autonomy	Does AI support or displace learner and teacher judgment?	Writing support, lesson planning, personalized pathways, automated feedback.	Professional development, appropriate-reliance policies, reflective use guidelines.
Academic integrity	How should authorship, assistance, assessment, and disclosure be redefined?	Generative AI for writing, coding, feedback, translation, and tutoring.	Assessment redesign, disclosure norms, process evidence, calibrated use of detection tools.

2.6 Methodological limitations:

This review is presented as a critical thematic synthesis rather than a quantitative meta-analysis. It does not estimate effect sizes and does not claim exhaustive coverage of all global literature. As noted in the reporting addendum (Appendix A), final database-export counts, duplicate-removal logs, and reasons for exclusion should be confirmed by the corresponding author against the original search records prior to journal submission to ensure full PRISMA 2020 compliance.

3. Conceptual Foundations: From Trustworthy AI to Human-Centered Education

Research on the ethics of AI in education is informed by broader frameworks of trustworthy and responsible AI. The Organisation for Economic Co-operation and Development (OECD, 2019/2024) frames trustworthy AI in education as a dual challenge: harnessing AI to strengthen educational processes while simultaneously preparing learners for societies increasingly shaped by automation. UNESCO's policy guidance similarly adopts a human-centered perspective grounded in rights, equity, safety, and the public good (Miao & Holmes, 2023). At the regulatory level, the European Union's Artificial Intelligence Act (European Parliament and Council of the European Union, 2024) classifies many educational AI applications as high-risk and imposes corresponding documentation, oversight, and risk-management obligations. These macro-level frameworks supply a shared vocabulary of principles, including fairness, transparency, robustness, accountability, and human oversight. However, the educational literature increasingly argues that such principles must be translated into domain-specific norms (Fu & Weng, 2024).

The transition from general AI ethics to educational ethics is consequential because education is not merely a service sector. Educational institutions are entrusted with developmental aims, civic formation, credentialing, and distributive justice. Mouta et al. (2024) argue that AI in education inherits many concerns associated with general AI ethics, but that these concerns become more acute when they affect learners' identities, opportunities, and relationships. From this perspective, AI does not simply enter education; it reconfigures how institutions understand and act upon students (Lockwood & Brown, 2024).

A notable development in recent scholarship is the shift toward human-centered and education-centered approaches. Fu and Weng (2024) argue that responsible AI in education should be conceptualized in terms of stakeholder relationships and characteristics, such as agency, equity, and intelligibility, rather than through technical performance alone. This reframing matters because many educational applications are justified on the basis of efficiency, while their ethical consequences typically arise in relational and pedagogical dimensions that technical metrics cannot adequately capture (Buragohain & Chaudhary, 2025).

4. Privacy, Datafication, and Surveillance

Privacy remains a foundational ethical concern in the discourse surrounding AI in education. AI systems in educational settings typically depend on large volumes of behavioral, performance, biometric, and interactional data. Learning analytics platforms, adaptive systems, online proctoring software, and generative AI assistants can collectively generate extensive records of student activity (Kohnke & Zaugg, 2025; Yambal & Waykar, 2025). Reviews by Fu and Weng (2024) and Yan et al. (2025) identify privacy and data security as foundational ethical concerns across K–20 educational settings. UNESCO's guidance on generative AI in education and research further cautions that these technologies pose distinctive risks, as user inputs and uploaded content may expose sensitive personal or academic information to external platforms (Miao & Holmes, 2023; Xue, Chinapah, & Zhu, 2025).

The literature further distinguishes between privacy as data protection and privacy as freedom from pervasive surveillance. The former concerns issues of consent, data retention, storage, third-party access, and lawful processing (Javed, 2024). The latter concerns the normalization of surveillance within educational environments. A student may formally consent to data processing yet still find themselves in a learning environment where their writing behavior, attendance, keystrokes, or emotional states are monitored in intrusive ways (Silva, Godwin, & Jayanagara, 2024). This broader concern with surveillance is particularly evident in research on automated proctoring, predictive analytics, and emotion recognition technologies, where the boundary between educational support and behavioral control becomes increasingly indistinct.

A persistent shortcoming in the literature is that privacy is frequently invoked as a principle rather than examined as a contextual institutional practice. Although many studies advocate for privacy-preserving design, few examine how educational institutions navigate trade-offs among student support, data minimization, procurement arrangements, and legal compliance (Feretzakis et al., 2024). The issue is therefore not merely the absence of adequate safeguards; educational institutions are also becoming increasingly dependent on data-intensive platforms whose business models and technical infrastructures lie outside pedagogical governance. As a result, students' educational data may be governed by opaque terms of service rather than by norms rooted in educational values (Jose, 2024).

The most compelling recent policy literature connects privacy to underlying power imbalances. UNESCO and OECD analyses note that AI may exacerbate disparities when vulnerable students are subjected to more intensive data extraction or are less able to understand and contest the systems that affect them (Miao & Holmes, 2023; Slimi & Villarejo Carballido, 2023). This point is ethically significant because students are often in a weak position to refuse AI-mediated systems without incurring academic disadvantage. Privacy in education is therefore not merely a matter of notice and consent; it also concerns whether learners exercise meaningful control over how they are computationally represented (Igbokwe, 2024).

5. Fairness, Bias, and Inclusion

Scholarship on AI in education increasingly emphasizes concerns regarding fairness and bias. These technologies have the potential to reproduce or amplify inequalities through multiple mechanisms, including training data composition, proxy variables, interface design, language assumptions, and decision-making procedures (Selvam & González Vallejo, 2025). Research in this area has progressed from the broad assertion that "algorithms can be biased" toward a more nuanced understanding that educational disparities can arise at multiple stages, including technology access, dataset representation, modeling decisions, output interpretation, and intervention implementation (Mishra et al., 2024).

Approaches that link algorithmic bias to broader social and educational frameworks offer particular analytical value. They emphasize disparities in access, representation, algorithmic decision-making, interpretation, and citizenship (Lockwood & Brown, 2024). This approach is analytically more comprehensive than narrow fairness metrics, as it recognizes that inequities may persist even when an algorithm appears statistically balanced, particularly if certain student groups are underrepresented in the design process, subjected to differential surveillance, or targeted by more intensive interventions. Fu and Weng (2024) similarly identify fairness and equity as essential components of responsible AI in education, while UNESCO and OECD documents emphasize inclusion, disability access, and cultural responsiveness (Miao & Holmes, 2023).

The literature on equity further highlights a tension between the promise of personalization and the substantive demands of inclusion. Adaptive learning systems are frequently characterized as individualized and equitable because they adjust content to student performance (Mahmoud & Sørensen, 2024). However, the literature increasingly asks: individualized for whom, and according to which conception of learning? An AI system may personalize the pace of instruction while still embedding normative assumptions about language, behavior, disability, or academic success (Kohnke & Zaugg, 2025). Recent OECD analyses are particularly pertinent here, warning that AI may exacerbate disparities unless systems are culturally responsive, accessible, and supported by substantial teacher development.

Recent reviews have consistently underscored the global-north and English-language bias of the current evidence base. Yan et al. (2025) and Maluleke (2025) both highlight the limited geographic diversity of the literature. This pattern is critical because fairness concerns are shaped by local infrastructures, languages, curricular traditions, and resource constraints. An ethical discourse derived predominantly from anglophone, higher-education, and well-resourced contexts risks treating context-specific assumptions as universal. Although the field is increasingly aware of this issue, comparative studies remain scarce (Mustofa, Kola, & Owusu-Darko, 2025; Zhang, 2025).

6. Transparency, Explainability, and Intelligibility

Opacity is a pervasive concern in the ethical discourse on educational AI, as stakeholders are frequently required to trust systems they do not fully understand. Transparency concerns arise in at

least three domains (Bulut & Beiting-Parrish, 2024). First, students and educators may be unaware that AI is being deployed, what data it relies upon, or how its outputs are generated (Rane, 2024). Second, even when the use of AI is disclosed, the rationale for its recommendations or scores may remain obscure to those lacking specialized knowledge. Third, institutions may not fully understand vendor systems, particularly when these tools are procured as black-box services (De Fine Licht & De Fine Licht, 2020).

Fu and Weng (2024) emphasize transparency and intelligibility as essential characteristics of responsible AI in education. Yan et al. (2025) and Chaudhry, Cukurova, and Luckin (2022) similarly note that explainability is a recurring concern across both empirical and theoretical research. In educational contexts, however, the demand for explainability extends beyond knowledge acquisition. It is intricately connected to fairness, due process, and pedagogical practice. Students must comprehend feedback in order to engage productively with the learning process (Lee & Moore, 2024). Educators require an understanding of the system's rationale to exercise professional judgment rather than passively accept machine-generated outputs. Administrators, in turn, need sufficient transparency to justify the institutional deployment of these systems and to responsibly adjudicate complaints (Bogina et al., 2021).

A discernible tendency in the academic literature is to treat explainability as a technical add-on rather than as an integral component of the educational process. Although a model may offer saliency maps, probability scores, or dashboard indicators, it may still fail to address the critical question in educational contexts: why should this recommendation be trusted in this specific situation? Educational transparency must therefore be both communicative and contestable, rather than merely comprehensible from a computational standpoint. Without this shift, explainability risks becoming a compliance exercise that does little to substantively empower learners or educators (Bura, Jonnalagadda, & Naayini, 2025; Gardezi et al., 2023).

The diffusion of generative AI has intensified this concern. Large language models are capable of producing coherent and persuasive explanations; however, the appearance of explanation does not equate to genuine transparency (Kadaruddin, 2023). UNESCO's guidance underscores that generative AI systems can produce content that is plausible but inaccurate, complicating their application in feedback, tutoring, and assessment (Miao & Holmes, 2023). Ethical concerns therefore extend beyond whether a system can articulate its reasoning to whether stakeholders can reliably distinguish substantiated explanations from persuasive simulations (Kotsis, 2025).

7. Accountability, Governance, and Institutional Responsibility

Accountability is where ethical principles intersect with institutional practice. The literature consistently observes that responsibility for AI in education is distributed across multiple actors (Ab Rahman et al., 2024). Developers design systems, administrators procure them, educators deploy them, and students experience their effects. Consequently, identifying responsibility can be difficult

when problems emerge. This distributed character helps explain why accountability is so frequently emphasized in review studies and policy recommendations (Yan et al., 2025).

Yan et al. (2025) identify ethical governance as a pivotal area for solutions in the field. Recent comparative analyses of university policies have shown that, while institutions have generally taken a proactive stance, they have not comprehensively addressed the challenges introduced by generative AI (Alali & Wardat, 2024; Adamakis & Rachiotis, 2025). Universities are increasingly formulating guidelines on acceptable use, academic integrity, and pedagogical innovation; however, these policies frequently lack specificity, internal consistency, or robust links to procurement practices and assessment design (Song, 2024). The literature suggests that many institutions have been faster to regulate student behavior than to govern third-party systems, data storage arrangements, procurement criteria, or to conduct impact assessments (Valerio, 2024).

The ethical significance of this imbalance is considerable. Institutions find it more straightforward to regulate students' use of AI than to establish internal accountability for their own AI adoption, although the latter may have more profound implications (Adamakis & Rachiotis, 2025). When AI tools are incorporated into grading, advising, writing support, or risk assessment, ethical responsibility cannot be devolved to individual users. Governance must therefore include clear documentation of objectives, role allocation, audit processes, appeal mechanisms, and substantive human oversight (Li, 2024). While the OECD AI Principles (OECD, 2019/2024) and UNESCO guidelines (Miao & Holmes, 2023) advocate for this approach, the education-specific literature indicates that implementation remains uneven (Vatankhah et al., 2024).

Research on governance has also emphasized the importance of participatory accountability. Ethical frameworks are more effective when students, educators, and affected communities are actively involved in the design and oversight of AI systems. Mouta et al. (2024) argue that ethical analysis should regard AI as a non-neutral artifact and treat its arrival as an opportunity to reassess broader educational ethics. This perspective is essential because accountability concerns not only the remediation of harms after they occur but also the question of who possesses the authority to define educational values before such systems are implemented (Thurzo, 2025).

8. Agency, Autonomy, and the Roles of Teachers and Learners

A notable strength of recent literature is its increasing emphasis on agency and autonomy. Ethical considerations in education extend beyond the prevention of harm to encompass the protection of learners' and educators' capacities to make informed decisions (Berson, Berson, & Luo, 2025). Fu and Weng (2024) specifically identify agency and autonomy as essential characteristics of responsible, human-centered AI. Guidelines from the OECD (2019/2024) and UNESCO (Miao & Holmes, 2023) similarly emphasize that AI should augment, rather than supplant, human decision-making.

In practice, however, agency can erode in subtle ways. Educators may feel pressured to conform to AI-generated recommendations or to rely on automated feedback under workload and performance

constraints (Duan & Zhao, 2024). Students may begin to treat generative AI as a substitute for drafting, reflection, or sustained engagement with complex material. In both cases, the central concern is not outright automation but the gradual erosion of skills and shifts in the norms that govern educational practice (Chun et al., 2025). Nikolic et al. (2024) similarly note that adoption of AI by teaching academics is advancing more rapidly than the professional learning required to support its responsible use.

The literature considers autonomy from two interrelated perspectives. For students, the question is whether AI facilitates self-directed learning or instead promotes dependency and uncritical acceptance (Peng & Li, 2025; Wu et al., 2024). For educators, the question is whether AI expands pedagogical approaches or constrains professional judgment through templates, predictions, and default settings. These questions are ethical in nature because education aims not only at efficient task performance but also at cultivating individuals capable of reasoning, deliberation, and engagement (Wang, Cui, & Yuan, 2025). A system that improves immediate task completion while undermining these capacities may be educationally detrimental, even where narrower performance metrics improve.

Recent empirical work on perceptions of generative AI challenges overly simplistic narratives. Both faculty and students typically acknowledge significant potential alongside notable risks in university courses (Jauhiainen & Guerra, 2023). Students' perspectives on AI in assessment are more nuanced than a binary of acceptance or rejection (Routray & Khandelwal, 2024). These findings suggest that agency should not be construed as a simple choice between using or avoiding AI. Rather, it involves negotiated norms regarding disclosure, appropriate reliance, evaluation, and responsibility (Igbokwe, 2024).

9. Academic Integrity, Assessment, and Authentic Learning

Since the emergence of widely available generative AI, no ethical issue has attracted greater public attention than academic integrity. The literature increasingly suggests, however, that discussions of integrity should extend beyond detection and prohibition (Costa, Mfolo, & Ntsobi, 2024). Generative AI blurs traditional boundaries between assistance and authorship, drafting and outsourcing, and editing and original creation (Wirzal et al., 2024). The ethical question is therefore not solely whether students engage in misconduct, but also how assessment should be designed for environments in which fluent machine-generated text is readily accessible (Rane et al., 2024).

Research on institutional policy supports this reorientation. Comparative analyses indicate that universities worldwide tend to prioritize academic integrity in their guidelines on generative AI, often placing it above detailed guidance on pedagogy or governance (Song, 2024; Slimi & Villarejo Carballido, 2023). While this emphasis is understandable, it is not sufficient. Detection tools do not provide a reliable solution: Weber-Wulff et al. (2023) demonstrated that commonly used AI-text detectors lack reliability and can be circumvented through paraphrasing, obfuscation, and similar

techniques. Excessive reliance on detection may therefore foster false confidence and risk unwarranted accusations.

Recent investigations of student behavior add further complexity. Chan (2024) and related work reveal that a substantial proportion of students do not perceive AI-assisted work as constituting traditional plagiarism. Students assess the ethics of AI in assessment through a tripartite lens that encompasses system design, institutional policy, and personal learning experience (Vieriu & Petrea, 2025). This perspective challenges narratives focused exclusively on misconduct by emphasizing that ethical use cannot be secured solely through monitoring or honor codes when students lack clear, shared, and educationally grounded norms.

Recent scholarship increasingly endorses assessment design that is process-oriented, authentic, and reflective. Such approaches include phased drafting, oral defenses, in-class practical applications, source documentation, and transparent disclosure of AI use (Lin & Chen, 2024). Their value lies not in being "AI-proof"—a standard that is likely unattainable—but in realigning assessment with learning processes that retain educational significance (Kalinichenko & Griban, 2025). Within this framing, the discourse on academic integrity points to a broader ethical question: whether the integration of AI will prompt institutions to develop more meaningful forms of assessment or, conversely, intensify surveillance-based responses (Özer, 2024).

10. Cross-Cutting Tensions

Several overarching tensions recur across themes in the literature. The first concerns personalization versus surveillance: many AI tools offer tailored support but require extensive data collection (Ünlü, 2024). The second concerns efficiency versus professional judgment: while automation can enhance efficiency and broaden access to feedback, it may also compel educators and institutions to prioritize scalable outcomes over thoughtful pedagogical practice (Pashaie, Mohammadi, & Golmohammadi, 2024). The third concerns innovation versus fairness: tools claimed to democratize education may inadvertently perpetuate disparities related to language, disability, culture, or infrastructure (Sánchez et al., 2025). The fourth concerns assistance versus integrity: generative AI can scaffold the organization of ideas, yet it may also obscure authorship and reduce meaningful cognitive engagement (Ghose, Ali, & Mishra, 2024). The fifth concerns distributed use versus concentrated responsibility: although AI affects numerous stakeholders, accountability for its negative effects often remains ambiguous (Slimi & Villarejo Carballido, 2023).

These tensions indicate that the central ethical question in AI and education is not whether the technology is inherently beneficial or detrimental. Rather, it concerns whether institutions can govern AI in a manner consistent with educational values. The literature is most productive when it eschews technological determinism and examines how ethical outcomes are shaped by purpose, context, design, and power dynamics. It is least productive when it presumes that broad principles or technical solutions can resolve contested educational issues (Feldstein, 2023).

Table (3). Cross-cutting ethical tensions and implications for practice

Tension	Why it matters	Implication for practice
Personalization vs. surveillance	Personalized support typically requires extensive behavioral and performance data.	Adopt data minimization, learner control mechanisms, and privacy-preserving analytics.
Efficiency vs. pedagogical autonomy	Automation may scale feedback but narrow the scope of teacher judgment.	Retain educator responsibility for consequential judgments and invest in AI literacy training.
Innovation vs. fairness	New tools may expand access while reproducing language, disability, or infrastructure inequalities.	Conduct contextual fairness and accessibility audits prior to deployment.
Automation vs. accountability	Responsibility becomes diffuse when vendors, administrators, and educators share control.	Define accountability, escalation, and appeal procedures in institutional policy.
Assistance vs. integrity	Generative AI can scaffold learning or substitute for meaningful engagement.	Redesign assessment around process, reflection, disclosure, and authentic performance.

11. Major Research Gaps

Although the literature on AI ethics in education has expanded rapidly, several substantial gaps remain.

First, the empirical evidence base lags behind normative debate. Numerous reviews observe that the field relies predominantly on conceptual papers, policy analyses, and stakeholder perception surveys. There is a relative paucity of longitudinal, comparative, or mixed-method studies that document how ethical risks and benefits manifest over time in classrooms, schools, and universities (Yan et al., 2025).

Second, higher education continues to dominate scholarly attention. Wieczorek et al. (2025) note that ethical analysis of AI in K–12 education has only recently begun to be systematically consolidated. This imbalance is significant because the ethical stakes, legal constraints, and learner vulnerabilities vary substantially across age groups.

Third, the scope of research is geographically narrow. Recent reviews consistently highlight the predominance of scholarship from Global North and English-language contexts (Maluleke, 2025). Substantially more research is needed in multilingual, lower-resource, and differently regulated educational environments.

Fourth, distinct technologies are too often treated as interchangeable. Ethical analyses frequently refer to “AI in education” as though predictive analytics, automated scoring, intelligent tutoring systems, recommender systems, and large language models raise identical concerns. This assumption is unwarranted. Technology- and use-case-specific analyses are therefore necessary.

Fifth, stakeholder engagement remains insufficiently developed. Although students are frequently positioned as the focal point of ethical consideration, they are seldom involved as co-creators of

ethical frameworks. Research on participatory design and governance is particularly limited in K–12 settings (Berson et al., 2025).

Sixth, procurement and platform governance have received limited scholarly attention. Institutional discussions of ethics frequently emphasize classroom applications while neglecting vendor contracts, data-sharing agreements, model update cycles, platform lock-in, and the political economy of educational AI.

Seventh, the field requires more robust measures for assessing educational harm and benefit. Ethical discussions often highlight potential risks without offering defensible methods for evaluating trade-offs. Further research is needed to evaluate not only fairness and privacy metrics but also pedagogical quality, epistemic agency, trust, and the distribution of educational opportunity.

12. Directions for Future Research and Practice

Future work should advance along four principal lines. First, ethical frameworks should become more context-sensitive, distinguishing among educational stages, disciplinary fields, technology types, and institutional objectives. Evaluating an AI writing assistant in a first-year composition course raises different considerations from evaluating a dropout-prediction model or a chatbot deployed in a K–12 setting.

Second, the field should strengthen its methodological rigor through longitudinal designs, implementation studies, and comparative case analyses. A stronger empirical foundation would help distinguish which ethical concerns remain primarily theoretical from those that manifest in practice, and would clarify how mitigation strategies perform in real educational settings.

Third, governance research should shift toward a more participatory and organizational orientation. Rather than focusing solely on student use of AI, scholars should examine how institutions procure AI systems, train staff, formulate policies, conduct technology audits, and establish mechanisms for resolving disputes and grievances. Policy documents acquire practical significance only through enactment.

Fourth, educational ethics must remain anchored in substantive educational purposes. The central question is not whether AI is integrated but how AI-mediated environments can advance learning, inclusion, critical thinking, and democratic participation. This requires evaluating AI not only as a productivity tool but also as a significant factor shaping what counts as knowledge, evidence, authorship, and participation in educational contexts.

13. Conclusion

Scholarship on the ethical dimensions of AI in education has established a relatively coherent foundation, with privacy, fairness, transparency, accountability, autonomy, and integrity emerging as central concerns. Recent reviews demonstrate increasing conceptual sophistication and clearer differentiation among stakeholder groups and educational levels. However, the field remains

fragmented and continues to privilege normative declarations over empirical evaluation of ethical concerns in practice.

The ethical considerations surrounding AI in education are distinctive because they implicate pedagogical and institutional dimensions, not only technical performance. AI systems shape how students are perceived, how educators make decisions, how assessment is organized, and how educational opportunities are distributed. Responsible implementation in educational contexts therefore cannot be reduced to compliance with general principles. It requires governance strategies that foreground human and educational values, protect learners, support educators, and preserve the centrality of educational purposes.

The most significant gap in the current literature is therefore both practical and theoretical: the field requires more robust evidence on how ethical frameworks are implemented in educational settings without reducing ethics to mere compliance. Until such research is more fully developed, institutions should adopt AI with caution, with participatory oversight, and with a clear recognition that education is not simply another domain for automated optimization.

References

- Ab Rahman, N. F., Wang, S. L., Ng, T. F., & Ghoneim, A. S. (2024). Artificial intelligence in education: A systematic review of machine learning for predicting student performance. Journal of Advanced Research in Applied Sciences and Engineering Technology, 54(1), 198–221. https://doi.org/10.37934/araset.54.1.198221
- Adamakis, M., & Rachiotis, T. (2025). Artificial intelligence in higher education: A state-of-the-art overview of pedagogical integrity, artificial intelligence literacy, and policy integration. Encyclopedia, 5(4), 180. https://doi.org/10.3390/encyclopedia5040180
- Alali, R., & Wardat, Y. (2024). Opportunities and challenges of integrating generative artificial intelligence in education. International Journal of Religion, 5(7), 784–793. https://doi.org/10.61707/8y29gv34
- Berson, I. R., Berson, M. J., & Luo, W. (2025). Innovating responsibly: Ethical considerations for AI in early childhood education. AI, Brain and Child, 1(1). https://doi.org/10.1007/s44436-025-00003-5
- Bogina, V., Hartman, A., Kuflik, T., & Shulner-Tal, A. (2021). Educating software and AI stakeholders about algorithmic fairness, accountability, transparency and ethics. International Journal of Artificial Intelligence in Education, 32(3), 808–833. https://doi.org/10.1007/s40593-021-00248-0
- Bulut, O., & Beiting-Parrish, M. (2024). The rise of artificial intelligence in educational measurement: Opportunities and ethical challenges. Chinese/English Journal of Educational Measurement and Evaluation, 5(3). https://doi.org/10.59863/miql7785

- Bura, C., Jonnalagadda, A. K., & Naayini, P. (2025). The role of explainable AI (XAI) in trust and adoption. Journal of Artificial Intelligence General Science (JAIGS), 7(1), 262–277. https://doi.org/10.60087/jaigs.v7i01.331
- Buragohain, D., & Chaudhary, S. (2025). Navigating ChatGPT in ASEAN higher education: Ethical and pedagogical perspectives. Computer Applications in Engineering Education, 33(4). https://doi.org/10.1002/cae.70062
- Chan, C. K. Y. (2024). Students' perceptions of 'AI-giarism': Investigating changes in understandings of academic misconduct. Education and Information Technologies, 30(6), 8087–8108. https://doi.org/10.1007/s10639-024-13151-7
- Chaudhry, M. A., Cukurova, M., & Luckin, R. (2022). A transparency index framework for AI in education. Center for Open Science. https://doi.org/10.35542/osf.io/bstcf
- Chun, J., Kim, J., Kim, H., Lee, G., Cho, S., Kim, C., Chung, Y., & Heo, S. (2025). A comparative analysis of on-device AI-driven, self-regulated learning and traditional pedagogy in university health sciences education. Applied Sciences, 15(4), 1815. https://doi.org/10.3390/app15041815
- Costa, K., Mfolo, L. N., & Ntsobi, M. P. (2024). Challenges, benefits and recommendations for using generative artificial intelligence in academic writing—A case of ChatGPT. Center for Open Science. https://doi.org/10.31222/osf.io/7hr5v
- De Fine Licht, K., & De Fine Licht, J. (2020). Artificial intelligence, transparency, and public decision-making. AI & Society, 35(4), 917–926. https://doi.org/10.1007/s00146-020-00960-w
- Duan, H., & Zhao, W. (2024). The effects of educational artificial intelligence-powered applications on teachers' perceived autonomy, professional development for online teaching, and digital burnout. International Review of Research in Open and Distributed Learning, 25(3), 57–76. https://doi.org/10.19173/irrodl.v25i3.7659
- Eden, C., Chisom, O., & Adeniyi, I. (2024). Integrating AI in education: Opportunities, challenges, and ethical considerations. Magna Scientia Advanced Research and Reviews, 10(2), 6–13. https://doi.org/10.30574/msarr.2024.10.2.0039
- European Parliament and Council of the European Union. (2024). Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). Official Journal of the European Union.
- Feldstein, S. (2023). Evaluating Europe's push to enact AI regulations: How will this influence global norms? Democratization, 31(5), 1049–1066.

  https://doi.org/10.1080/13510347.2023.2196068

- Feretzakis, G., Papaspyridis, K., Gkoulalas-Divanis, A., & Verykios, V. S. (2024). Privacy-preserving techniques in generative AI and large language models: A narrative review. Information, 15(11), 697. https://doi.org/10.3390/info15110697
- Fu, Y., & Weng, Z. (2024). Navigating the ethical terrain of AI in education: A systematic review on framing responsible human-centered AI practices. Computers and Education: Artificial Intelligence, 7, 100306. https://doi.org/10.1016/j.caeai.2024.100306
- Gardezi, M., Joshi, B., Rizzo, D. M., Ryan, M., Prutzer, E., Brugler, S., & Dadkhah, A. (2023). Artificial intelligence: A useful analogy for explainable AI. Agronomy Journal, 116(3), 1217–1228. https://doi.org/10.1002/agj2.21353
- Ghose, A., Ali, S. M. A., & Mishra, A. M. (2024). Artistic expressions, generative AI, and legal tapestry (pp. 198–214). IGI Global. https://doi.org/10.4018/979-8-3693-1950-5.ch011
- Gore, S., & Dove, E. (2024). Ethical considerations in the use of artificial intelligence in counselling and psychotherapy training: A student stakeholder perspective—A pilot study. Counselling and Psychotherapy Research, 25(1). https://doi.org/10.1002/capr.12770
- Igbokwe, I. C. (2024). Artificial intelligence in educational leadership: Risks and responsibilities. European Journal of Arts, Humanities and Social Sciences, 1(6), 3–10.

  https://doi.org/10.59324/ejahss.2024.1(6).01
- Jauhiainen, J. S., & Guerra, A. G. (2023). Generative AI and ChatGPT in school children's education: Evidence from a school lesson. Sustainability, 15(18), 14025.

  https://doi.org/10.3390/su151814025
- Javed, F. (2024). The evolution of artificial intelligence in teaching and learning of English language in higher education: Challenges, risks, and ethical considerations (pp. 249–276). Emerald. https://doi.org/10.1108/978-1-83549-486-820241015
- Jose, D. (2024). Data privacy and security concerns in AI-integrated educational platforms. Recent Trends in Management and Commerce, 5(2), 87–91. https://doi.org/10.46632/rmc/5/2/19
- Kadaruddin, K. (2023). Empowering education through generative AI: Innovative instructional strategies for tomorrow's learners. International Journal of Business, Law, and Education, 4(2), 618–625. https://doi.org/10.56442/ijble.v4i2.215
- Kalinichenko, O., & Griban, G. (2025). Theoretical foundations of using artificial intelligence technologies in physical education of high school students. Scientific Journal of National Pedagogical Dragomanov University, Series 15, 11(198), 74–77. https://doi.org/10.31392/udu-nc.series15.2025.11(198).14

- Kohnke, S., & Zaugg, T. (2025). Artificial intelligence: An untapped opportunity for equity and access in STEM education. Education Sciences, 15(1), 68.

  https://doi.org/10.3390/educsci15010068

- Kotsis, K. T. (2025). From potential to practice: Rethinking STEM education through artificial intelligence. International Journal of Advanced Multidisciplinary Research and Studies, 5(5), 613–617. https://doi.org/10.62225/2583049x.2025.5.5.4972

- Lee, S. S., & Moore, R. L. (2024). Harnessing generative AI (GenAI) for automated feedback in higher education: A systematic review. Online Learning, 28(3).

  https://doi.org/10.24059/olj.v28i3.4593

- Li, Z. (2024). AI ethics and transparency in operations management: How governance mechanisms can reduce data bias and privacy risks. Journal of Applied Economics and Policy Studies, 13(1), 89–93. https://doi.org/10.54254/2977-5701/13/2024130

- Lin, H., & Chen, Q. (2024). Artificial intelligence (AI)-integrated educational applications and college students' creativity and academic emotions: Students' and teachers' perceptions and attitudes. BMC Psychology, 12(1). https://doi.org/10.1186/s40359-024-01979-0

- Lockwood, A., & Brown, J. (2024). Mitigating AI bias in school psychology: Toward equitable and ethical implementation. Center for Open Science. https://doi.org/10.31234/osf.io/mh4rj

- Mahmoud, C. F., & Sørensen, J. T. (2024). Artificial intelligence in personalized learning with a focus on current developments and future prospects. Research and Advances in Education, 3(8), 25–31. https://doi.org/10.56397/rae.2024.08.04

- Maluleke, A. F. (2025). AI adoption in African higher education: A systematic review of benefits and ethical implications. Interdisciplinary Journal of Education Research, 7(2), a05. https://doi.org/10.38140/ijer-2025.vol7.2.05

- Miao, F., & Holmes, W. (2023). Guidance for generative AI in education and research. UNESCO.

- Mishra, I., Kashyap, V., Yadav, N., & Pahwa, D. (2024). Harmonizing intelligence: A holistic approach to bias mitigation in artificial intelligence (AI). International Research Journal on Advanced Engineering Hub, 2(7), 1978–1985. https://doi.org/10.47392/irjaeh.2024.0270

- Mouta, A., Pinto-Llorente, A. M., & Torrecilla-Sánchez, E. M. (2024). Uncovering blind spots in education ethics: Insights from a systematic literature review on artificial intelligence in education. International Journal of Artificial Intelligence in Education, 34, 1166–1205. https://doi.org/10.1007/s40593-023-00384-9

- Mustofa, H. A., Kola, A. J., & Owusu-Darko, I. (2025). Integration of artificial intelligence (ChatGPT) into science teaching and learning. International Journal of Ethnoscience and Technology in Education, 2(1), 108. https://doi.org/10.33394/ijete.v2i1.14195
- Nikolic, S., Wentworth, I., Sheridan, L., Moss, S., Duursma, E., Jones, R. A., Ros, M., & Middleton, R. (2024). A systematic literature review of attitudes, intentions and behaviours of teaching academics pertaining to AI and generative AI (GenAI) in higher education: An analysis of GenAI adoption using the UTAUT framework. Australasian Journal of Educational Technology. https://doi.org/10.14742/ajet.9643
- Organisation for Economic Co-operation and Development (OECD). (2019, updated 2024). Recommendation of the Council on Artificial Intelligence. OECD Legal Instruments.
- Özer, M. (2024). Potential benefits and risks of artificial intelligence in education. Bartın Üniversitesi Eğitim Fakültesi Dergisi, 13(2), 232–244. https://doi.org/10.14686/buefad.1416087
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., et al. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. BMJ, 372, n71. https://doi.org/10.1136/bmj.n71
- Pashaie, S., Mohammadi, S., & Golmohammadi, H. (2024). Unlocking athlete potential: The evolution of coaching strategies through artificial intelligence. Proceedings of the Institution of Mechanical Engineers, Part P: Journal of Sports Engineering and Technology. https://doi.org/10.1177/17543371241300889
- Peng, J., & Li, Y. (2025). Frontiers of artificial intelligence for personalized learning in higher education: A systematic review of leading articles. Applied Sciences, 15(18), 10096. https://doi.org/10.3390/app151810096
- Rane, N. (2024). Enhancing the quality of teaching and learning through ChatGPT and similar large language models: Challenges, future prospects, and ethical considerations in education. TESOL and Technology Studies, 5(1), 1–6. https://doi.org/10.48185/tts.v5i1.1000
- Rane, N., Shirke, S., Choudhary, S. P., & Rane, J. (2024). Education strategies for promoting academic integrity in the era of artificial intelligence and ChatGPT: Ethical considerations, challenges, policies, and future directions. Journal of ELT Studies, 1(1).

  https://doi.org/10.48185/jes.v1i1.1314
- Routray, R., & Khandelwal, K. (2024). Artificial intelligence (AI) adoption: Do Generation Z students feel technostress in deploying AI for completing courses of study at universities? Asian Education and Development Studies, 13(5), 534–545. https://doi.org/10.1108/aeds-06-2024-0115

- Sánchez, E., Aguilar, P., Hidalgo, A., Ramírez, A., Hidalgo, J., & Guamán, E. (2025). Artificial intelligence for inclusive classrooms transforming learning through innovation and digital equity. Revista Multidisciplinar de Estudios Generales, 4(4), 595–615. https://doi.org/10.70577/reg.v4i4.322
- Schönberger, M. (2024). Integrating artificial intelligence in higher education: Enhancing interactive learning experiences and student engagement through ChatGPT (pp. 11–34). Emerald. https://doi.org/10.1108/978-1-83549-486-820241002
- Selvam, M., & González Vallejo, R. (2025). Ethical and privacy considerations in AI-driven language learning. LatIA, 3, 328. https://doi.org/10.62486/latia2025328
- Silva, G., Godwin, G., & Jayanagara, O. (2024). The impact of AI on personalized learning and educational analytics. International Transactions on Education Technology, 3(1), 36–46. https://doi.org/10.33050/itee.v3i1.669
- Slimi, Z., & Villarejo Carballido, B. (2023). Navigating the ethical challenges of artificial intelligence in higher education: An analysis of seven global AI ethics policies. TEM Journal, 12(2), 590–602. https://doi.org/10.18421/tem122-02
- Song, N. (2024). Higher education crisis: Academic misconduct with generative AI. Journal of Contingencies and Crisis Management, 32(1). https://doi.org/10.1111/1468-5973.12532
- Taşkın, M. (2025). Artificial intelligence in personalized education: Enhancing learning outcomes through adaptive technologies and data-driven insights. Human Computer Interaction, 8(1), 173. https://doi.org/10.62802/ygye0506
- Thurzo, A. (2025). Provable AI ethics and explainability in next-generation medical and educational AI agents: Trustworthy ethical firewall. MDPI AG. https://doi.org/10.20944/preprints202502.2232.v1
- Ünlü, S. C. (2024). Enhancing user experience through AI-driven personalization in user interfaces. Human Computer Interaction, 8(1), 19. https://doi.org/10.62802/m7mqmb52
- Valerio, A. S. (2024). Anticipating the impact of artificial intelligence in higher education: Student awareness and ethical concerns in Zamboanga City, Philippines. Cognizance Journal of Multidisciplinary Studies, 4(6), 408–418. https://doi.org/10.47760/cognizance.2024.v04i06.024
- Vatankhah, S., Bamshad, V., Arici, H. E., & Duan, Y. (2024). Ethical implementation of artificial intelligence in the service industries. The Service Industries Journal, 44(9–10), 661–685. https://doi.org/10.1080/02642069.2024.2359077
- Vieriu, A. M., & Petrea, G. (2025). The impact of artificial intelligence (AI) on students' academic development. Education Sciences, 15(3), 343. https://doi.org/10.3390/educsci15030343

- Wang, K., Cui, W., & Yuan, X. (2025). Artificial intelligence in higher education: The impact of need satisfaction on artificial intelligence literacy mediated by self-regulated learning strategies. Behavioral Sciences, 15(2), 165. https://doi.org/10.3390/bs15020165
- Weber-Wulff, D., Anohina-Naumeca, A., Bjelobaba, S., Foltýnek, T., Guerrero-Dib, J., Popoola, O., Šigut, P., & Waddington, L. (2023). Testing of detection tools for AI-generated text. International Journal for Educational Integrity, 19(1). https://doi.org/10.1007/s40979-023-00146-z
- Wieczorek, M., Hosseini, M., & Gordijn, B. (2025). Unpacking the ethics of using AI in primary and secondary education: A systematic literature review. AI and Ethics, 5, 4693–4711. https://doi.org/10.1007/s43681-025-00770-0
- Wirzal, M. D. H., Md Nordin, N. A. H., Abd Halim, N. S., & Bustam, M. A. (2024). Generative AI in science education: A learning revolution or a threat to academic integrity? A bibliometric analysis. Jurnal Penelitian dan Pengkajian Ilmu Pendidikan: E-Saintika, 8(3), 319–351. https://doi.org/10.36312/e-saintika.v8i3.2127
- Wu, D., Zhang, S., Ma, Z., Yue, X.-G., & Dong, R. K. (2024). Unlocking potential: Key factors shaping undergraduate self-directed learning in AI-enhanced educational environments. Systems, 12(9), 332. https://doi.org/10.3390/systems12090332
- Xue, Y., Chinapah, V., & Zhu, C. (2025). A comparative analysis of AI privacy concerns in higher education: News coverage in China and Western countries. Education Sciences, 15(6), 650. https://doi.org/10.3390/educsci15060650
- Yambal, S., & Waykar, Y. A. (2025). Future of education using adaptive AI, intelligent systems, and ethical challenges (pp. 171–202). IGI Global. https://doi.org/10.4018/979-8-3693-6527-4.ch006
- Yan, Y., Liu, H., & Chau, T. (2025). A systematic review of AI ethics in education: Challenges, policy gaps, and future directions. Journal of Global Information Management, 33(1), 1–50. https://doi.org/10.4018/JGIM.386381
- Zhang, X. (2025). Fairness and effectiveness in AI-driven educational assessments: Challenges and mitigation strategies. Journal of Innovation and Development, 11(1), 7–10.

  https://doi.org/10.54097/adcxze34

Appendix A. PRISMA 2020 Reporting Addendum

This addendum documents the reporting details required for full PRISMA 2020 compliance. Items that can be specified at the protocol level are completed; items requiring the corresponding author's final search records are flagged for verification prior to submission.

A1. Information sources and search strategy:

Element	Specification and author actions
Databases searched	Scopus, Web of Science, ERIC, Education Source, IEEE Xplore, ACM Digital Library, ScienceDirect, and SpringerLink. Confirm the final list of databases actually searched.
Supplementary sources	UNESCO, OECD, European Union, and national education-policy repositories, supplemented by forward and backward citation chasing of key reviews. Confirm exact sources used.
Date range	Primary emphasis on 2023–2025 literature, with foundational and policy sources retained where required. Confirm exact database search dates.
Core search string	("artificial intelligence" OR "AI" OR "generative AI" OR "large language model" OR "learning analytics" OR "adaptive learning") AND (education OR school OR university OR "higher education" OR "K–12") AND (ethics OR privacy OR fairness OR bias OR transparency OR explainability OR accountability OR autonomy OR "academic integrity" OR inclusion OR governance).
Language limits	Specify any language restrictions applied during screening.
Protocol registration	If the review protocol was registered (e.g., PROSPERO, OSF), insert the registry name and identifier; if not registered, state: "No review protocol was registered."

A2. PRISMA flow information:

PRISMA stage	Count/status	Author verification required
Records identified from databases	[Insert exact number]	Required from database exports.
Records identified from policy sources and citation chasing	[Insert exact number]	Required from search log.
Duplicate records removed	[Insert exact number]	Required from reference manager or screening tool.
Records screened by title and abstract	[Insert exact number]	Required from screening spreadsheet.
Reports sought for retrieval	[Insert exact number]	Required from full-text screening log.
Reports not retrieved	[Insert exact number]	Required if any full texts were unavailable.
Reports assessed for eligibility	[Insert exact number]	Required from full-text screening log.
Reports excluded with reasons	[Insert by reason]	Suggested categories: not education-focused; not AI-focused; no ethical dimension; duplicate; non-scholarly or vendor source; full text unavailable.
Sources included in qualitative synthesis	[Insert final count]	Verify the final included corpus against the citations used in the synthesis.

A3. Data extraction fields:

Field	Definition
Bibliographic details	Author, year, title, publication venue, and DOI or URL.
Educational context	K–12, higher education, teacher education, early childhood, professional learning, or cross-sector.
AI technology/application	Generative AI, learning analytics, adaptive learning, automated assessment, intelligent tutoring system, recommender system, proctoring system, or governance/policy framework.
Stakeholder group	Students, teachers, faculty, administrators, institutions, policymakers, vendors, or affected communities.
Ethical domain	Privacy/surveillance, fairness/bias/inclusion, transparency/explainability, accountability/governance, agency/autonomy, academic integrity, or other.
Evidence type	Systematic review, meta-review, empirical study, conceptual paper, policy guidance, or technical and governance analysis.
Main finding	Concise statement of the principal finding relevant to AI ethics in education.
Limitations	Methodological, contextual, geographic, or evidentiary limitations reported by the source or identified during synthesis.

A4. Quality appraisal plan:

For full systematic-review submission, study-type-appropriate appraisal tools should be applied: AMSTAR 2 for systematic reviews, the Mixed Methods Appraisal Tool (MMAT) or the Joanna Briggs Institute tools for empirical studies, and a transparent authority–relevance–implementation checklist for policy documents. As the present manuscript does not report formal quality scores, its conclusions are framed as a qualitative synthesis of current scholarship rather than as a quantitative estimate of effect or prevalence.

A5. Declarations required prior to submission:

Declaration	Text to complete
Funding	[Insert funding source and grant number, or state: "This research received no external funding."]
Conflicts of interest	[Insert any conflicts, or state: "The author(s) declare no competing interests."]
Ethics approval	Not applicable, as this is a literature review involving no human participants or identifiable private data.
Data availability	[Insert statement, e.g., "The search strategy and screening log are available from the corresponding author upon reasonable request."]
Author contributions	[Insert contribution statement using the target journal's required taxonomy, such as CRediT.]
Use of AI tools	[Disclose any AI-assisted editing or drafting in accordance with the target journal's policy.]

Declarations

Funding. [To be completed by the corresponding author. If no funding was received, state: "The author(s) received no specific funding for this work."]

Conflicts of interest. [To be completed by the corresponding author. If none, state: "The author(s) declare no competing interests."]

Ethics approval. Not applicable. This article is a literature review and reports no research involving human participants, animals, or identifiable private data.

Data availability. No primary datasets were generated or analyzed beyond the published literature reviewed in this article. The search strategy and screening log are available from the corresponding author upon reasonable request.

Author contributions. [To be completed by the corresponding author in accordance with the target journal's contribution taxonomy, such as CRediT.]

Use of AI tools. [Disclose any AI-assisted editing or drafting in accordance with the target journal's policy.]