STM JOURNALS

International Journal of
Education Sciences

ISSN: 3048-9784
Volume 3, Issue 1, 2026
January–June
DOI (Journal): 10.37591/IJES

https://journals.stmjournals.com/ijes

Research IJES

Ethical Risks of Generative AI in Education: Challenges, Implications, and a Responsible Use Framework

Samyo Ranjan Jagdev[1,*], Chinmay Giri[2], G. Subhasmita Prusty[3], Girija Nandan Das[4]

Abstract

The rapid diffusion of generative artificial intelligence (AI) technologies in educational settings is reshaping how teaching, learning, and assessment are designed and enacted. Large language models and related generative systems offer powerful capabilities for content creation, personalized feedback, and instructional support, promising gains in efficiency and learner engagement. However, their growing use also introduces a complex set of ethical risks that challenge foundational educational values such as integrity, equity, transparency, and trust. This paper critically examines the ethical implications of generative AI in education through a comprehensive synthesis of existing literature and policy discussions. The analysis identifies six interrelated ethical risk dimensions that consistently emerge across educational contexts: academic integrity and authorship, cognitive dependency, bias, and inequality, privacy, and data protection, transparency, and accountability, and assessment fairness. Rather than treating these concerns in isolation, the study demonstrates how they interact and reinforce one another, amplifying potential negative impacts on learning quality, student outcomes, and institutional credibility. In particular, the opacity of generative AI systems and uneven access to AI tools raise significant concerns regarding fairness, explainability, and the legitimacy of assessment practices. Building on this analysis, the paper proposes a four-pillar ethical framework for the responsible integration of generative AI in education, grounded in integrity, equity, transparency, and privacy. The framework translates ethical principles into actionable institutional and pedagogical strategies, including assessment redesign, AI disclosure policies, bias auditing, AI literacy initiatives, and compliance with data protection regulations. By aligning ethical considerations with practical implementation, the framework offers a holistic approach that moves beyond purely technical solutions. The paper contributes to ongoing debates on AI ethics in education by consolidating fragmented ethical concerns into a unified structure and providing guidance for educators, administrators, and policymakers. It argues that the educational value of generative AI ultimately depends not on technological capability alone, but on the ethical frameworks that govern its use, ensuring innovation supports meaningful, equitable, and trustworthy learning experiences.

***Author for Correspondence**
Samyo Ranjan Jagdev
E-mail: samyojagdev@gmail.com

[1]Assistant Professor, Department of Computer Science Engineering, GIET, Bhubaneswar, Odisha, India
[2]Student, Department of Computer Science Engineering, GIET, Bhubaneswar, Odisha, India
[3]Student, Department of Computer Science Engineering, GIET, Bhubaneswar, Odisha, India
[4]Student, Department of Computer Science Engineering, GIET, Bhubaneswar, Odisha, India

Received Date: January 23, 2025
Accepted Date: January 28, 2025
Published Date: February 01, 2026

Citation: Samyo Ranjan Jagdev, Chinmay Giri, G. Subhasmita Prusty, Girija Nandan Das. Ethical Risks of Generative AI in Education: Challenges, Implications, And a Responsible Use Framework. Journal of Education Sciences. 2026; 3(1): 23-30p.

Keywords: Academic integrity, AI ethics, bias, and fairness, education technology, generative artificial intelligence, responsible AI

INTRODUCTION

The emergence of generative artificial intelligence (AI), particularly large language models (LLMs) such as GPT-based systems, has rapidly accelerated the integration of AI technologies into educational environments [1, 2]. These systems are capable of producing human-like text, solving complex problems, generating programming code, and offering personalized feedback, making them increasingly attractive and versatile tools for both students and educators. As a

result, generative AI is reshaping core educational practices, including instruction, assessment, feedback, and curriculum design shown in Figure 1 [3, 4].

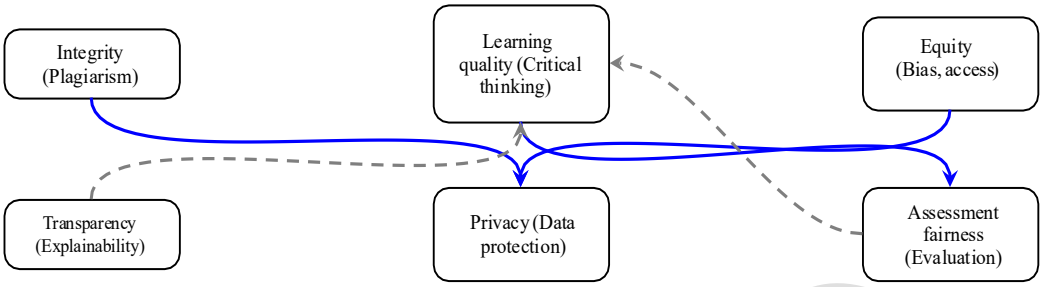

Figure 1. Research gaps in ethical studies of generative AI in education.

Students use generative AI tools for drafting essays, summarizing readings, practicing problem-solving, and preparing for examinations [5]. Educators, meanwhile, employ these systems for lesson planning, quiz generation, automated grading, and administrative support [6]. Proponents argue that generative AI can improve efficiency, support personalized learning, and reduce educators' workload, enabling greater focus on high-impact pedagogical activities [7].

Despite these advantages, the rapid and often unregulated adoption of generative AI raises significant ethical concerns. One of the most important challenges involves academic integrity. The ability of AI systems to generate high-quality essays and solutions complicates traditional notions of authorship and originality, making it increasingly difficult to distinguish between human- and AI-generated work samples [8, 9].

Cognitive dependency is another concern. Overuse of generative AI may discourage deep engagement, critical thinking, and metacognitive skill development, particularly when students use AI as a substitute rather than a supplement to learning [10, 11].

Equity and bias represent major ethical dimensions. Generative AI models are trained on large-scale datasets that may encode historical, cultural, and linguistic biases, leading to unequal or discriminatory outputs [12–14]. Privacy and data protection further complicate the ethical landscape. Generative AI systems often collect and process sensitive student data, including written assignments, interaction logs, and personal identifiers. Without having robust safeguards, such practices risk violating data protection regulations and eroding trust between students and institutions [15, 16]. Adherence to frameworks such as GDPR and FERPA is essential.

Finally, the opacity of many generative AI systems poses challenges for transparency and accountability. The "black-box" nature of these models makes it difficult for educators to explain or justify AI-assisted decisions related to grading, feedback, or student support [17, 18]. This lack of explainability undermines trust and complicates ethical responsibility when errors or biases occur.

In response to these challenges, this paper provides a comprehensive examination of ethical risks associated with generative AI in education and proposes a four-pillar ethical framework centered on integrity, equity, transparency, and privacy shown in (Figure 2).

BACKGROUND AND RELATED WORK

AI in education has a rich history, ranging from intelligent tutoring systems to automated grading and learning analytics [7, 1]. With the recent emergence of generative AI, research focus has shifted toward understanding the ethical implications of these technologies [3, 4].

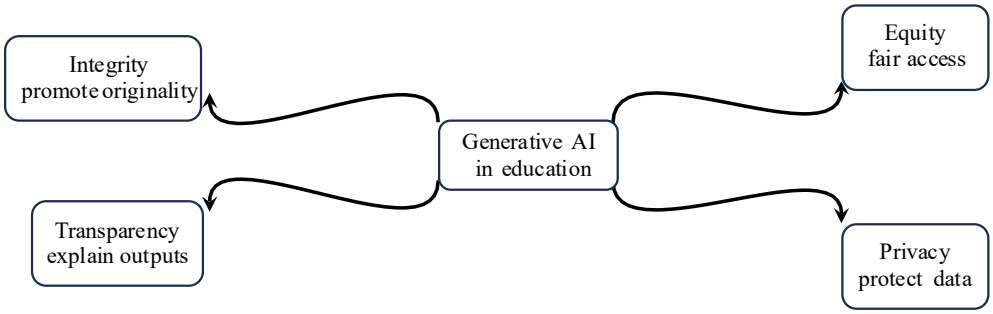

Figure 2. Four-Pillar Ethical Framework for responsible generative AI use in education.

Several studies have documented challenges, including difficulty in detecting AI-generated work, over-dependence on AI, privacy concerns related to student data, and biases embedded in AI models [12, 13, 15, 16]. However, comprehensive frameworks addressing all dimensions of AI ethics in education remain limited.

Research Gap

Key research gaps include:

- *Fragmented focus:* Studies often examine one ethical issue at a time
- *Lack of practical frameworks:* Few frameworks guide ethical implementation
- *Limited empirical data:* Student and educator perceptions are underexplored
- *Equity and accessibility:* Ensuring fair outcomes for all students is rarely studied
- *Transparency and accountability:* AI decisions in grading and feedback are poorly understood

ETHICAL RISKS OF GENERATIVE AI

Academic Integrity and Authorship

Generative AI can produce essays, reports, and coding assignments. Institutions must consider AI-detection tools, revised honor codes, and clear policies on AI usage [8].

Cognitive Dependency

Over-reliance on AI can reduce development of critical thinking, research, and independent reasoning skills [11].

Bias and Inequality

AI models reflect biases in training data, which can disadvantage certain student groups. Bias audits and inclusive datasets are key mitigation strategies [12, 13].

Privacy and Data Protection

Sensitive student data may be collected and processed. Institutions must comply with GDPR and FERPA to protect student privacy [15, 16].

Transparency and Accountability

AI often operates as a "black box," complicating grading and feedback. Explainable AI methods are necessary [17, 18].

Results

Although this study does not involve empirical data collection, it yields several important analytical and conceptual results derived from a systematic synthesis of prior literature and ethical analyses. The results clarify the nature of ethical risks posed by generative AI in education and demonstrate the value of a unified ethical framework for responsible adoption.

Identification of Core Ethical Risk Dimensions

The first result of this study is the identification and consolidation of six core ethical risk dimensions associated with generative AI in education: academic integrity, cognitive dependency, bias, and inequality, privacy, and data protection, transparency, and accountability, and assessment fairness. While prior studies often examine these issues in isolation, this work demonstrates that these risks consistently emerge across diverse educational contexts and stakeholder groups.

Importantly, the analysis shows that these risks are not independent. For example, lack of transparency exacerbates fairness concerns in assessment, while bias in AI models can intensify equity gaps when access to AI tools is uneven. This interdependence highlights the inadequacy of fragmented policy responses.

Mapping Ethical Risks to Educational Impacts

A second key result is the systematic mapping of ethical risks to concrete educational impacts, as summarized in Table 1. The findings indicate that ethical risks translate directly into measurable educational consequences, such as reduced originality in student work, degradation of higher-order thinking skills, and erosion of trust in institutional assessment systems.

Table 1. Ethical risks and educational impacts.

Dimension	Key Risk	Impact
Integrity	AI-generated assignments	Reduced originality
Learning Quality	Cognitive dependency	Skill degradation
Equity	Bias and access inequality	Unequal outcomes
Privacy	Data misuse	Loss of trust
Transparency	Black-box decisions	Reduced accountability
Fairness	AI-assisted cheating	Invalid assessment

This mapping clarifies how ethical challenges move beyond abstract concerns and materially affect learning quality, student outcomes, and institutional credibility. In particular, assessment fairness emerges as a critical pressure point, where multiple ethical dimensions converge.

Development of a Four-Pillar Ethical Framework

This study proposes a four-pillar ethical framework – Integrity, Equity, Transparency, and Privacy – to guide the responsible integration of generative artificial intelligence in educational contexts. The framework is theoretically grounded in established principles from educational ethics, responsible AI governance, and socio-technical systems theory. Rather than treating ethical challenges as isolated technical issues, the framework conceptualizes generative AI as a socio-educational system whose impacts emerge from interactions among technology, institutional policy, pedagogical design, and human agency.

The four pillars function as interdependent normative foundations that collectively regulate the ethical use of generative AI across teaching, learning, and assessment. Each pillar addresses multiple ethical risk dimensions identified in the literature while remaining flexible enough to accommodate diverse educational contexts.

Integrity: Preserving Academic Honesty and Cognitive Engagement

The integrity pillar is grounded in traditional principles of academic ethics, including honesty, originality, and intellectual responsibility. In the context of generative AI, integrity extends beyond plagiarism prevention to encompass cognitive integrity, defined as meaningful learner engagement in the intellectual processes underlying knowledge construction.

Generative AI challenges conventional notions of authorship by enabling students to outsource cognitive labor such as idea generation, argument formulation, and problem-solving. While such tools can support learning when used reflectively, unregulated use risks eroding essential academic skills, including critical thinking, synthesis, and metacognition. From a constructivist learning perspective, learning occurs through active cognitive effort; excessive reliance on AI-generated outputs undermines this process.

The integrity pillar therefore emphasizes:

- Clear disclosure of AI assistance in academic work
- Assessment designs that prioritize reasoning, reflection, and process over final output
- Pedagogical practices that frame AI as a cognitive scaffold rather than a substitute for learnin
- By integrating integrity into both policy and pedagogy, this pillar reframes academic honesty as a developmental goal rather than solely a compliance issue.

Equity: **Ensuring Fairness, Inclusion, and Equal Opportunity**

The equity pillar is rooted in theories of social justice, inclusive education, and algorithmic fairness. It addresses the risk that generative AI may exacerbate existing educational inequalities due to unequal access, differential digital literacy, and embedded algorithmic bias.

Generative AI systems are trained on large-scale datasets that reflect historical and cultural biases, potentially producing outputs that disadvantage marginalized groups. Additionally, unequal access to advanced AI tools can confer disproportionate advantages to students with greater financial, linguistic, or technological resources. These dynamics conflict with the educational principle of equal opportunity.

This pillar conceptualizes equity as both distributive and procedural:

- Distributive equity involves ensuring equitable access to AI tools, infrastructure, and trainin
- Procedural equity requires ongoing bias audits, inclusive dataset evaluation, and attention to how AI-mediated decisions affect different student population
- By embedding equity into institutional AI governance, this pillar shifts responsibility from individual users to systemic safeguards that promote fairness across diverse learner groups.

Transparency: Enabling Explainability, Accountability, and Trust

The transparency pillar is grounded in responsible AI principles and accountability theory. It addresses the ethical risks posed by the opaque or "black-box" nature of many generative AI systems, particularly when these tools influence assessment, feedback, or academic decision-making.

In educational settings, opacity undermines trust by preventing students and educators from understanding how AI-generated outputs are produced or how AI-assisted decisions are made. This lack of explainability complicates accountability when errors, bias, or unfair outcomes occur.

Transparency within this framework operates at three levels:

- *Technical transparency:* clarity about system capabilities, limitations, and training data source
- *Pedagogical transparency:* explicit communication about when and how AI is used in teaching, assessment, and feedbac
- *Institutional transparency:* documented policies outlining responsibility, oversight, and avenues for contesting AI-influenced decision
- By promoting explainability and human-in-the-loop oversight, the transparency pillar reinforces ethical accountability and sustains institutional trust in AI-supported educational practices.

Privacy: Protecting Student Data and Informational Autonomy

The privacy pillar is grounded in data ethics, informational self-determination, and regulatory frameworks such as GDPR and FERPA. Generative AI systems frequently process sensitive educational

data, including written assignments, behavioral interaction logs, and personal identifiers. Without adequate safeguards, such practices risk unauthorized data use, surveillance, and erosion of student autonomy.

This pillar conceptualizes privacy not merely as regulatory compliance but as a foundational ethical right that supports learner trust and academic freedom.

It emphasizes:

- Data minimization and purpose limitatio
- Secure storage and responsible data governanc
- Informed consent and clear communication about data usag

By foregrounding privacy, the framework acknowledges that ethical AI adoption depends on protecting students' control over their personal and intellectual data.

Interdependence of the Four Pillars

A key theoretical contribution of the framework lies in its recognition of the interdependence among the four pillars. Integrity is undermined when transparency is lacking; equity is compromised when privacy protections are uneven; accountability is impossible without explainability. Ethical governance of generative AI therefore requires a holistic approach rather than isolated technical or policy interventions. This interdependent structure positions the framework as a dynamic ethical system capable of adapting to evolving AI capabilities and educational practices. It also aligns with socio-technical perspectives that emphasize the co-evolution of technology, institutions, and human values.

Actionable Alignment Between Ethics and Practice

Another important result is the translation of ethical principles into actionable recommendations, presented in Table 2. The framework demonstrates that ethical AI adoption is not solely a technical challenge but an organizational and pedagogical one. Clear disclosure policies, AI literacy training, bias audits, and data protection compliance emerge as feasible and immediately applicable interventions.

Table 2. Ethical framework and recommended actions.

Pillar	Recommended Actions
Integrity redesign Equity	AI disclosure policies, assessment Equal AI access, bias audits
Transparency	AI literacy training
Privacy	Data protection compliance

This alignment between ethical values and institutional practices represents a concrete contribution of the study, bridging the gap between ethical theory and educational implementation.

Clarification of Research and Policy Gaps

Finally, the results highlight persistent gaps in current research and policy approaches. The analysis reveals limited empirical evaluation of ethical frameworks, insufficient attention to student perspectives, and a lack of cross-cultural considerations in AI ethics research. These gaps underscore the need for future empirical validation and policy experimentation.

Overall, the results demonstrate that responsible integration of generative AI in education requires a holistic ethical approach rather than isolated technical fixes.

Assessment Fairness

AI-assisted work may confer unfair advantages. Assessments should be redesigned to integrate AI responsibly [9].

DISCUSSION

Ethical risks are interconnected. Addressing one dimension without considering others can lead to incomplete solutions. The four-pillar framework enables holistic responsible adoption.

Policy and Pedagogical Implications

- Redesign assessments to emphasize critical thinking
- Introduce AI ethics and literacy curricula
- Establish clear institutional AI guidelines
- Encourage human-in-the-loop oversight for AI-assisted evaluations

Limitations and Future Work

- Conceptual study; empirical validation needed
- Surveys and interviews with diverse populations
- Cross-cultural studies of AI ethics
- Evaluate framework effectiveness in practice

CONCLUSION

The rapid integration of generative artificial intelligence into educational environments represents a profound shift in how teaching, learning, and assessment are conducted. While these technologies offer substantial opportunities for personalization, efficiency, and pedagogical innovation, this study demonstrates that their uncritical adoption poses significant ethical risks that can undermine core educational values.

This paper contributes a comprehensive synthesis of ethical challenges associated with generative AI in education, consolidating previously fragmented concerns into six interrelated risk dimensions: academic integrity, cognitive dependency, bias, and inequality, privacy, and data protection, transparency, and accountability, and assessment fairness. The results show that these risks are deeply interconnected and that addressing them in isolation is insufficient. Ethical shortcomings in one dimension can rapidly propagate across others, amplifying negative educational outcomes.

A central contribution of this study is the proposed four-pillar ethical framework – integrity, equity, transparency, and privacy – which provides a structured and actionable approach for responsible generative AI adoption. Rather than treating ethics as an abstract or purely technical concern, the framework translates ethical principles into concrete institutional and pedagogical actions, including assessment redesign, AI disclosure policies, bias audits, AI literacy initiatives, and compliance with data protection regulations. This framework offers educators, administrators, and policymakers a practical tool to balance innovation with responsibility.

From a pedagogical perspective, the findings underscore the need to rethink assessment and instructional design in the presence of generative AI. Traditional evaluation methods that prioritize content reproduction are increasingly vulnerable to misuse, whereas assessments emphasizing critical thinking, reflection, and process-oriented learning are more resilient and educationally valuable. Embedding AI literacy and ethical reasoning into curricula is therefore essential to empower students to engage with AI tools responsibly and critically.

At the policy level, the study highlights the importance of clear institutional governance structures for AI use in education. Transparent guidelines, human-in-the-loop oversight, and accountability mechanisms are necessary to maintain trust in AI-assisted educational processes. Without such safeguards, the use of generative AI risks exacerbating existing inequalities and eroding confidence in educational assessment systems.

This study is not without limitations. As a conceptual analysis, it does not empirically evaluate the effectiveness of the proposed ethical framework. Future research should empirically test the framework

across diverse educational contexts, incorporate student and educator perspectives, and examine cross-cultural differences in ethical expectations and regulatory environments. Longitudinal studies assessing the impact of ethical AI integration on learning outcomes and equity are particularly needed.

In conclusion, generative AI should not be viewed as either a threat to education or a panacea for its challenges. Its educational value depends fundamentally on how it is governed, integrated, and used. By adopting a holistic ethical framework grounded in integrity, equity, transparency, and privacy, educational institutions can harness the benefits of generative AI while safeguarding the principles that underpin meaningful and equitable learning.

REFERENCES

1. Luckin R, Holmes W, Griffiths M, Forcier L. Intelligence Unleashed: An Argument for AI in Education. 1st edition. London, UK: Pearson; 2016. pp. 1–50.
2. Holmes W, Bialik M, Fadel C. Artificial Intelligence in Education: Promise and Implications. 1st edition. Boston, US: Educational Technology; 2019. pp. 1–210.
3. Zawacki-Richter O, Marín VI, Bond M, Gouverneur F. Systematic review of AI in higher education. International Journal of Educational Technology in Higher Education. 2021;18(1):1–25.
4. Selwyn N. Should Robots Replace Teachers?. 1st edition. Cambridge, UK: Polity Press; 2019. pp. 1–160. 10.1186/s41239–019-0171–0
5. Kasneci E, et al. ChatGPT for good?. Learning and Individual Differences. 2023;103:102274. 10.1016/j.lindif.2023.102274
6. Holmes W, Bialik M, Fadel C. Ethics of AI in education. Computers and Education: Artificial Intelligence. 1st edition. New York, US: Elsevier; 2022. pp. 1–35.
7. Baker RS, Inventado PS. Educational data mining and learning analytics. In: Larusson JA, White B, editors. Learning Analytics. 1st edition. New York, US: Springer; 2014. pp. 61–94.
8. Cotton D, et al. ChatGPT and academic integrity. Assessment & Evaluation in Higher Education. 2023;48(1):1–15.
9. McGee P. Academic integrity in the age of AI. Educational Technology Research. 2023;71(2):1–12.
10. Brynjolfsson E, Rock D, Syverson C. Artificial intelligence and the modern productivity paradox. Journal of Economic Perspectives. 2017;31(2):33–58.
11. Doran J. Cognitive offloading and AI tools in learning. Computers in Human Behavior. 2023;140:107561.
12. Bolukbasi T, et al. Man is to computer programmer as woman is to homemaker?. Advances in Neural Information Processing Systems. 2016;29:1–9.
13. Mehrabi N, Morstatter F, Saxena N, Lerman K, Galstyan A. A survey on bias and fairness in machine learning. ACM Computing Surveys. 2021;54(6):1–35.
14. Selwyn N. Digital education and inequality. British Journal of Sociology of Education. 2020;41(1):1–14.
15. Pardo A, Siemens G. Ethical dimensions of learning analytics. British Journal of Educational Technology. 2019;45(3):438–450.
16. Regan PM, Jesse J. Ethical challenges of edtech, big data and learning analytics. EDUCAUSE Review. 2019;54(3):1–10.
17. Doshi-Velez F, Kim B. Towards a rigorous science of interpretable machine learning. arXiv preprint arXiv:1702.08608. 2017.
18. Arrieta AB, et al. Explainable Artificial Intelligence (XAI). Information Fusion. 2020;58:82–115.

