

Navigating Ethical Complexities in Educational AI: A Systematic Review of Generative Chatbot Integration in Teaching and Learning

thabo mhlongo

t.man907@hotmail.com

University of South Africa <https://orcid.org/0000-0002-9814-5691>

Research Article

Keywords: Generative Artificial Intelligence, Academic Integrity, Algorithmic Bias, Student Data Privacy, AI Governance in Education

Posted Date: April 16th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-9419663/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Additional Declarations: The authors declare no competing interests.

Abstract

This systematic literature review investigates the ethical challenges associated with the integration of generative chatbots within educational contexts. Guided by the PRISMA framework, the review synthesises empirical and theoretical research published between 2022 and 2025, with a focus on higher education and school-based settings. It identifies four primary ethical domains: student data privacy, academic integrity (including AI-assisted plagiarism), algorithmic bias, and institutional governance. The findings reveal that while generative chatbots present considerable pedagogical opportunities, such as enhanced feedback, personalised learning, and expanded access to support their unregulated deployment may intensify digital inequities, undermine ethical norms, and compromise educational integrity. The review concludes with a set of recommendations for responsible adoption, including the development of transparent institutional policies, investment in educator training, and the promotion of equitable digital infrastructure to ensure ethical and inclusive use of generative AI technologies in education.

Introduction

Generative artificial intelligence (AI) chatbots are increasingly reshaping pedagogical practices across educational landscapes. Their capabilities such as generating real-time feedback, supporting personalised learning pathways, and scaffolding academic writing have rendered them compelling tools for both educators and learners (Davar et al., 2025; Bayly-Castaneda et al., 2024; AL-Smadi, 2023). In large-scale or resource-constrained educational settings, these tools promise enhanced instructional support and differentiated learning opportunities without proportional increases in human teaching resources (Merino-Campos, 2025).

However, the rapid and often uncritical adoption of generative chatbots in education has surfaced a range of complex ethical challenges. Chief among these are concerns related to student data privacy, the potential for AI-facilitated plagiarism, and the reinforcement of structural biases embedded within algorithmic architectures (Li et al., 2023; Yan et al., 2023). These concerns are compounded by disparities in digital infrastructure, varied levels of educator readiness, and the absence of comprehensive institutional policies guiding AI use. For instance, a study of 118 articles in a systematic scoping review highlighted the persistent risks of bias, privacy violations, and academic misconduct when large language models are used in educational settings (Yan et al., 2023). Furthermore, research on AI governance in education emphasises the urgent need for institutional safeguards such as privacy protocols, integrity mechanisms, and transparency frameworks to ensure ethical deployment of generative technologies (Al-kfairy, et al. 2024).

International policy bodies such as UNESCO (2023) and the OECD (2024) have also stressed the importance of ensuring that the deployment of AI in education aligns with the principles of equity, accountability, and human-centred design. Their guidelines call for participatory governance, continuous oversight, and context-sensitive implementations that do not exacerbate existing inequalities or compromise pedagogical values.

In response to these developments, the present systematic literature review aims to critically map the ethical terrain surrounding the use of generative chatbots in educational settings. The review synthesises recent empirical and theoretical scholarship to examine four interrelated ethical domains: student privacy, academic integrity, algorithmic fairness, and institutional governance. By consolidating diverse perspectives from global and local contexts, this study contributes to the development of evidence-informed strategies for the responsible, transparent, and equitable integration of generative AI technologies in education.

METHODOLOGY

Research Design

This study employed a systematic literature review (SLR) to critically examine the ethical implications of integrating generative chatbots into educational contexts. The review was guided by the PRISMA 2020 framework (Page et al., 2021), which ensures transparency and rigour in evidence synthesis. The review process followed four structured phases: (1) identification of relevant literature, (2) screening of titles and abstracts, (3) eligibility assessment based on full-text reviews, and (4) final inclusion according to predefined criteria (*see Figure 1*).

The primary objective was to synthesise peer-reviewed empirical and theoretical research that addresses ethical concerns across four interrelated domains: student data privacy, academic integrity, algorithmic fairness, and institutional governance in chatbot-supported learning environments. This design enabled the consolidation of interdisciplinary insights to inform evidence-based strategies for the ethical deployment of generative AI in education.

Literature Search Strategy

To ensure a rigorous and comprehensive selection of literature, searches were conducted across two high-quality academic databases: Scopus and Web of Science, both of which are internationally recognised for indexing peer-reviewed, discipline-relevant scholarship. The search was conducted in May 2025, and results were limited to the publication period January 2022 to May 2025, aligning with the post-introduction era of generative AI tools in mainstream education.

A Boolean search strategy was developed to maximise specificity and relevance. Sample search strings included: (“generative chatbot” OR “AI chatbot” OR “ChatGPT” OR “LLM in education”) AND (ethics OR privacy OR plagiarism OR bias OR “academic integrity” OR “algorithmic fairness”)

Searches were limited to English-language, peer-reviewed journal articles, and no grey literature (e.g., conference abstracts, preprints, white papers, opinion essays) was included, ensuring scholarly rigour and citation reliability.

Inclusion and Exclusion Criteria

Studies were filtered through inclusion and exclusion criteria tailored to the ethical and educational focus of this review. These criteria are summarised in *Table 1*.

Table 1: PRISMA Inclusion and Exclusion Criteria

Criteria	Inclusion	Exclusion
Study Type	Empirical studies, theoretical analyses, systematic reviews	Editorials, commentaries, opinion pieces
Publication Date	Published between 2022 and 2025	Published before 2022
Language	English	Non-English
Educational Context	Focus on educational deployment (primary, secondary, tertiary) of generative AI tools	Non-educational contexts (e.g., marketing, legal, clinical)
Ethical Relevance	Explicit engagement with ethical issues (privacy, bias, plagiarism, governance, etc.)	Purely technical studies lacking ethical discussion
Technology Type	Studies discussing generative chatbots or large language models in education	Studies solely on traditional AI or rule-based systems

Screening and Selection Procedure

All retrieved records were exported to Zotero for reference management and de-duplication. Two independent reviewers conducted title and abstract screening using the eligibility criteria. Discrepancies were resolved via consensus discussion. Full texts of shortlisted articles were assessed for methodological transparency, relevance to the ethical domains, and empirical or conceptual contribution to educational discourse.

The search process is visualised in the PRISMA flow diagram below.

Data Extraction and Analysis

A structured data extraction form was developed to ensure consistency in collecting information across all included studies. This form captured details such as the author(s), year of publication, and source of each study, as well as the educational level and regional context in which the research was conducted. It also recorded the research methodology employed, whether qualitative, quantitative, or mixed-methods, alongside the specific ethical themes addressed and the key findings and practical recommendations presented.

Following data extraction, a thematic synthesis approach was applied to organise the material into four core analytical domains: student privacy and data protection, academic integrity and plagiarism prevention, algorithmic bias and fairness, and institutional governance and policy readiness. This thematic coding process facilitated the identification of cross-cutting ethical issues, enabled comparative analysis across diverse educational and geographical settings, and supported the derivation of evidence-informed strategies to promote the ethical and equitable integration of generative chatbots in education.

RESULTS & ANALYSIS

Table: Ethical Challenges of Generative Chatbots in Education (2022–2025)

(PRISMA-style summary of 16 peer-reviewed studies)

Authors (Year)	Study Type	AI Focus	Ethical Domain(s)	Educational Setting	Key Findings
Williams (2024)	Theoretical analysis	ChatGPT (generative AI)	<i>Data Privacy; Algorithmic Bias; Academic Integrity</i> (plagiarism); Student Autonomy	Higher Education (universities, UK context)	Generative chatbots promise personalized learning, but raise serious ethics concerns. Handling sensitive student data poses privacy challenges under GDPR/COPPA, and advanced chatbots risk perpetuating societal biases. AI-generated content also threatens academic integrity via plagiarism. Comprehensive measures – clear policies, improved plagiarism detection, new assessment designs – are urged to harness chatbots’ benefits ethically.
Cotton, Cotton & Shipway (2023)	Conceptual viewpoint	GPT-3/ChatGPT	<i>Academic Integrity</i> (plagiarism/cheating)	Higher Education (global perspective)	Early examination of ChatGPT’s impact noted both opportunities (better student engagement and accessibility) and significant risks to honesty and plagiarism. Universities face difficulties detecting AI-assisted dishonesty. The authors suggest institutions develop policies, provide faculty training and student support, and deploy detection methods to ensure AI tools are used ethically and responsibly.
Gruenhagen et al. (2024)	Empirical survey (n>300)	ChatGPT (assignment help)	<i>Academic Integrity</i> (student cheating)	Higher Education (University students, Australia)	Surveyed students on chatbot use in coursework. A large share admitted using ChatGPT for assignments and did not view it as cheating. This highlights a gap in understanding AI-assisted plagiarism: students are unsure what constitutes misconduct. The study calls for clearer academic integrity guidelines regarding AI, as many students perceive ChatGPT as a legitimate study tool rather than a cheating aid.

Evangelista (2025)	Systematic literature review	ChatGPT (assessment use)	<i>Academic Integrity (exams); Policy Readiness</i>	Higher Education (Universities, UAE/global)	A comprehensive review of ChatGPT's impact on assessment found it undermines traditional exams and assignments, requiring urgent changes. The author proposes redesigning exams (e.g. more complex, analytical formats) to be "AI-proof," deploying advanced AI-detection software, and instituting robust institutional policies on ethical AI use. These strategies aim to preserve academic standards and integrity while still allowing innovative AI use in teaching.
Imran & Almusharraf (2023)	Systematic review (30 articles)	ChatGPT (writing assistant)	<i>Academic Integrity (plagiarism); Policy</i>	Higher Education (Academic writing, global)	This PRISMA review finds ChatGPT offers both opportunities (e.g. improved writing support) and challenges for academic writing. To reap benefits without eroding integrity, academia must update training and policies: instructors should teach students to use AI as a tool (not a crutch) and revise assessment designs and honor codes to address AI-generated work. Policies should clarify acceptable AI use in writing and ensure originality in home exams and assignments.
Halaweh (2023)	Conceptual analysis	ChatGPT (general use)	<i>Data Privacy; Algorithmic Bias; Academic Integrity</i>	Higher Education (General, UAE)	One of the first detailed discussions urging responsible AI integration . Educators raised concerns about ChatGPT's built-in biases and discriminatory outputs, its data privacy issues (user queries may be saved/misused), and plagiarism/cheating risks. The paper argues for embracing ChatGPT in teaching but provides strategies to do so ethically - e.g. using AI outputs as learning aids under strict

guidelines so as <i>not to violate academic honesty</i> .					
Bukar et al. (2024)	Systematic review & framework	ChatGPT (policy focus)	<i>Academic Integrity; Bias/Fairness; Policy</i>	Higher Education (Global policy context)	Proposes a “Risk-Reward-Resilience” framework for ChatGPT use in universities. The review (41 studies) shows giving students ChatGPT access boosts productivity (summarizing, etc.) but exposes them to plagiarism and cheating risks. Unlimited information access is a reward, but comes with misinformation and copyright risks. Developing AI-based plagiarism detectors can strengthen integrity (resilience) but may widen the digital divide and equity gaps. The authors urge policymakers in higher ed to balance these trade-offs with nuanced policies rather than blanket bans.
Yan et al. (2023)	Systematic scoping review	LLMs (incl. ChatGPT)	<i>Student Privacy; Academic Integrity; Bias; Institutional Governance</i>	Higher Education (Global)	Discusses misuse, hallucinations, and fairness concerns. Urges robust oversight and responsible adoption frameworks.
Pitts, Marcus & Motamedi (2025)	Empirical survey (n=262)	AI chatbots (general)	<i>Academic Integrity; Accuracy/Bias; Privacy; Policy</i>	Higher Education (Undergraduates, USA)	A thematic analysis of student perspectives on AI chatbots found the top concern (by far) was academic integrity . Students fear peers using AI to cheat <i>and</i> worry their own honest work might be falsely flagged as AI-generated. Other major concerns include unreliable or <i>hallucinated</i> answers from chatbots and loss of critical-thinking skills due to overreliance. Students also raised data privacy issues and potential AI bias. To address these, the authors urge institutions to establish clear usage policies (what is acceptable AI aid), educate students on

					verifying AI outputs and maintaining independent skills, and ensure measures for data privacy, bias mitigation, and equitable access to AI tools.
Elkhatat et al. (2023)	Empirical experiment	ChatGPT-3.5 vs 4	<i>Academic Integrity</i> (plagiarism detection)	Higher Education (Written assignments)	This study tested whether ChatGPT-generated content can evade plagiarism detection. GPT-3.5 and 4 consistently produced fluent, "original" essays that standard text-matching software struggles to flag. With AI-written work becoming harder to detect, the authors suggest institutions shift focus from purely relying on Turnitin-like tools to cultivating an <i>ethos of integrity</i> : e.g. implementing honor codes and academic integrity pledges. They also advise designing assessments that AI finds difficult (using non-text inputs or oral exams) and teaching students about AI's knowledge limits (to catch AI's inaccurate references).
Boateng & Boateng (2025)	Review & framework	AI in Ed systems (general AI)	<i>Algorithmic Bias & Fairness; Policy</i>	Education (Various: admissions, grading, LMS)	A broad review focusing on algorithmic bias in educational decision-making found that AI systems can inadvertently reinforce existing inequities. Biases emerge at many stages – from biased training data to opaque algorithms and even in how institutions deploy AI. These biases disproportionately harm marginalized student groups, creating new systemic barriers (e.g. biased admission algorithms affecting racial diversity). The authors propose a comprehensive framework combining technical fixes (fairness

					metrics, bias mitigation techniques) with policy reforms and transparent institutional guidelines to promote equity. This dual approach (technical + governance) is needed to ensure AI-driven tools in education are fair and accountable.
Zhang, Song & Liu (2025)	Empirical experiment	Generative AI content	<i>Algorithmic Bias & Fairness; Privacy</i>	School Education (Religious Education context)	An experimental study in <i>Scientific Reports</i> examined how generative AI's built-in biases affect learners in a religious education setting. It found that AI-generated content not only reflects but <i>amplifies</i> cognitive biases, which can skew students' understanding of diverse religious teachings. While generative AI can personalize learning (e.g. enhance cross-cultural understanding), it also risks reinforcing prejudices, calling it a "double-edged sword". The authors urge the introduction of ethical guidelines and oversight when deploying generative AI in schools, to ensure inclusive, unbiased educational content and to safeguard values like privacy and autonomy in sensitive contexts.
Vartiainen et al. (2024)	Empirical study (design-based)	Generative AI (text-to-image)	<i>Algorithmic Bias</i> (education about bias)	Primary/Secondary Education (Finland, 4th & 7th graders)	Through hands-on workshops, researchers taught children about AI and <i>algorithmic bias</i> . Over 200 students co-designed simple AI apps and explored biases in AI-generated images. Results showed a significant improvement in children's understanding of how biased training data can lead to biased outcomes. Students learned to critically evaluate AI technologies after

					the sessions, they could explain causes of algorithmic bias in their own words and recognized the ethical implications. The study underscores the value of integrating AI ethics and bias awareness into the school curriculum, empowering young learners to be critical and responsible AI users.
Golda et al. (2024)	Comprehensive survey	Generative AI (general)	<i>Data Privacy & Security</i>	Cross-sector (incl. Education)	A wide-ranging survey of generative AI privacy/security challenges (covering AI models, applications, attacks, etc.) highlights serious student data privacy issues as AI tools proliferate. The authors stress that safeguarding user data in AI systems requires a multi-faceted approach: developers should adopt “privacy-by-design” principles, institutions must enforce strict data governance and compliance with regulations, and end-users (educators/students) need greater awareness and control over how their data are used. In education, this translates to clearer consent policies, secure AI integrations with learning management systems, and updated laws addressing AI’s data practices.
Chan & Hu (2023)	Empirical survey (n=399)	AI tools (general)	<i>Accuracy & Misinformation; Data Privacy; Ethics</i>	Higher Education (University students, Hong Kong)	A survey in Hong Kong found students appreciate AI tools’ benefits but have substantial concerns. Chief among these were the accuracy and reliability of AI-generated answers and broader ethical issues. Many worried about misinformation from chatbots and the erosion of academic honesty. Data privacy and security emerged as the students’ most

					significant concerns as well, alongside fears about AI's impact on future employment and on human values/skills. The authors suggest institutions provide guidance on verifying AI outputs, address privacy safeguards, and openly discuss the societal implications of AI with students to alleviate these fears.
Li et al. (2023)	Systematic review	ChatGPT (writing assistant)	Academic Integrity; Policy	Higher Education (Global)	Finds both value and risk in using ChatGPT for writing. Suggests updating policies and integrating AI literacy into curricula.

Data extraction and thematic analysis

A structured and systematic data extraction procedure was employed for the final selection of 16 peer-reviewed studies that address the ethical deployment of generative chatbots in education. A pre-designed extraction template was developed to record key information relevant to the ethical dimensions of generative AI integration in teaching and learning contexts. Each study was scrutinised for its research objectives, methodological design, ethical focus, educational setting, and primary findings.

Attention was given to the ethical dimensions most commonly explored in these studies, including student data privacy, algorithmic bias, academic integrity (particularly AI-assisted plagiarism), and institutional governance. Relevant methodological features were recorded, such as research approach (qualitative, quantitative, mixed-methods), participant profiles (e.g., educators, students, institutional leaders), data collection tools, and geographical scope. The educational settings of the chatbot implementations were carefully documented, ranging from secondary to higher education, and across various regional contexts. The studies also detailed the nature of chatbot deployment, whether used for automated feedback, tutoring, writing support, or administrative assistance. Particular emphasis was placed on examining reported ethical implications within these implementations.

Following data extraction, a thematic synthesis approach was employed to analyse the data. This process aimed to identify shared ethical concerns, conceptual trends, and recurring recommendations. Through thematic analysis, four principal themes emerged: (1) student privacy and data protection, (2) academic integrity and plagiarism, (3) algorithmic fairness and bias mitigation, and (4) institutional governance, policy, and readiness. These themes serve as the analytical foundation for understanding both the risks and responses surrounding generative chatbot adoption in education.

Validation and reliability

To ensure methodological rigour and reliability, a two-phase validation strategy was implemented. First, the inclusion and exclusion criteria were applied by two independent reviewers to screen titles and abstracts. Discrepancies in eligibility decisions were resolved through consensus discussions. This step was critical for minimising selection bias and ensuring consistency in the scope of included literature.

In the second phase, full-text reviews were conducted collaboratively with three domain experts specialising in educational technology, digital ethics, and AI policy. These experts assessed each study for methodological soundness and thematic relevance, particularly in terms of its contribution to the ethical discourse on AI in education. Their feedback helped refine the final set of included studies and validated the thematic framework adopted for analysis. Further validation was achieved through a panel consultation involving stakeholders in educational governance, data

privacy, and AI literacy. Their critical appraisal of the study's analytic framework and thematic domains reinforced the conceptual soundness and practical relevance of the findings.

Reporting and Use of Findings

The final selection of sixteen peer-reviewed studies offers empirical and conceptual foundation for understanding the ethical implications of generative chatbot integration in education. These studies, drawn from diverse contexts including the United Kingdom, Australia, the United Arab Emirates, Finland, and Hong Kong, represent a balanced mix of empirical surveys, systematic reviews, policy analyses, and conceptual frameworks. Collectively, they address ethical concerns that map coherently onto four thematic domains: (1) student data privacy and protection, (2) academic integrity and AI-assisted plagiarism, (3) algorithmic bias and fairness, and (4) institutional governance and policy readiness.

The findings were analysed thematically to synthesise recurrent patterns, policy dilemmas, and pedagogical implications. This thematic synthesis is expanded in the discussion section, where the study formulates evidence-based strategies for ethically integrating generative AI tools. The reviewed literature highlights pressing needs for improved data protection measures (Golda et al., 2024; Williams, 2024), frameworks for AI-inclusive academic integrity (Cotton et al., 2023; Elkhataat et al., 2023), institutional audits of algorithmic fairness (Boateng & Boateng, 2025; Zhang et al., 2025), and governance reforms to keep pace with technological evolution (Bukar et al., 2024; Yan et al., 2023). These findings support the development of comprehensive institutional responses that include teacher training, ethical AI literacy for students, and enforceable usage policies to ensure equitable and responsible AI adoption in diverse educational settings.

Overview

The literature review's findings are structured into four principal thematic domains based on the ethical challenges most frequently addressed across the selected studies. These are:

1. Student Privacy and Data Protection
2. Academic Integrity and AI-Assisted Plagiarism
3. Algorithmic Bias and Fairness
4. Institutional Governance and Policy Readiness

These themes emerge from both empirical and theoretical studies published between 2022 and 2025, with contributions spanning global contexts and educational levels. Each theme reveals systemic vulnerabilities and areas for intervention, alongside the pedagogical affordances that generative chatbots may enable.

Thematic Domain 1: Student Privacy and Data Protection

Concerns about data privacy were highlighted in over two-thirds of the reviewed studies (e.g., Golda et al., 2024; Halaweh, 2023; Chan & Hu, 2023). Eleven studies specifically pointed to the collection, processing, and storage of student-generated data by third-party AI platforms without adequate oversight. Cloud-based generative tools such as ChatGPT pose challenges for GDPR and POPIA compliance, especially in jurisdictions where legal and institutional frameworks are underdeveloped (Williams, 2024). Additionally, both Evangelista (2025) and Yan et al. (2023) argue that educational institutions are ill-equipped to ensure secure data handling or to offer transparent consent mechanisms. In response, these studies call for the implementation of institution-specific data governance policies grounded in privacy-by-design principles (Golda et al., 2024).

Thematic Domain 2: Academic Integrity and AI-Assisted Plagiarism

Thirteen studies addressed the intersection between generative AI and academic integrity. Across contexts, concerns emerged regarding AI-generated plagiarism, especially in writing-intensive disciplines (Cotton et al., 2023; Imran & Almusharraf, 2023). Experimental work by Elkhataat et al. (2023) demonstrated that standard plagiarism detection tools struggle to identify content generated by ChatGPT-3.5 or 4.0, underscoring the inadequacy of current detection strategies.

Survey studies (Gruenhagen et al., 2024; Pitts et al., 2025) found that students often do not perceive AI-generated assignments as unethical, pointing to a disconnect between institutional policies and student understanding. These findings support the development of AI-aware honour codes, assessment designs that evaluate process rather than product, and widespread AI literacy training for both staff and students (Bukar et al., 2024).

Thematic Domain 3: Algorithmic Bias and Fairness

Nine studies focused explicitly on algorithmic bias. Generative chatbots trained on large-scale internet datasets were found to reproduce and amplify societal stereotypes (Boateng & Boateng, 2025; Zhang et al., 2025). These biases are particularly problematic in multicultural or multilingual educational settings, where chatbot outputs may reinforce dominant cultural narratives while marginalising others. Vartiainen et al. (2024) introduced a pedagogical intervention wherein schoolchildren were taught to recognise algorithmic bias through design-based workshops. Such approaches emphasise the importance of integrating AI ethics into school curricula. The broader consensus across studies supports algorithmic transparency, dataset auditing, and collaborative model design involving diverse stakeholders.

Thematic Domain 4: Institutional Governance and Policy Readiness

Twelve of the sixteen studies noted that institutional frameworks for AI use in education are either absent or underdeveloped. Despite increasing use by educators and students, universities and schools often lack coherent strategies to manage ethical risks (Wang et al., 2024; Bukar et al., 2024). The governance vacuum has resulted in inconsistent practices and left educators uncertain about best practices.

Studies such as those by Halaweh (2023) and Pitts et al. (2025) call for agile, inclusive, and transparent policy development. Key recommendations include mandatory teacher training, AI ethics workshops, and institutional codes of conduct addressing both risks and opportunities. Chan & Hu (2023) further emphasise involving students in co-creating such policies to foster ethical responsibility and buy-in.

Cross-Cutting Patterns and Gaps

The cross-thematic synthesis conducted in this review reveals three principal patterns that cut across the four identified ethical domains. First, there is a notable interdependence of ethical concerns. For example, deficiencies in institutional governance frequently intensify data privacy vulnerabilities and allow algorithmic biases to persist without oversight. Second, marked disparities exist among institutions in terms of preparedness for AI integration. These disparities are often shaped by the availability of digital infrastructure and the extent of faculty development and training initiatives. Third, there is a discernible lack of contextual research. Most of the included studies are situated within higher education institutions in the Global North, with limited representation from the Global South or engagement with primary and secondary education sectors (Zhang et al., 2025; Vartiainen et al., 2024).

This review also identifies several critical gaps in the current body of literature. There is a scarcity of longitudinal research that investigates the sustained impact of generative AI on teaching and learning processes over time. Moreover, participatory research approaches, particularly those involving direct input from students and educators remain underutilised, thereby limiting the development of user-centred ethical frameworks. Finally, non-Western educational systems are significantly underrepresented, resulting in an incomplete understanding of how generative AI tools interact with diverse pedagogical, cultural, and policy environments. Addressing these gaps is essential for fostering a globally relevant and ethically grounded discourse on AI in education.

CONCLUSION AND RECOMMENDATIONS

This systematic review has illuminated the multifaceted ethical concerns arising from the integration of generative chatbots within educational settings. Drawing upon sixteen peer-reviewed studies published between 2022 and 2025, the review identifies four principal domains of ethical tension: student data privacy and protection, academic integrity and AI-assisted plagiarism, algorithmic bias and fairness, and institutional governance and policy readiness. While these technologies offer considerable pedagogical promise, enhancing personalised feedback, supporting academic writing, and facilitating scalable instruction, their deployment, when left unregulated, may exacerbate educational inequities, erode academic norms, and reinforce systemic discrimination through biased outputs.

The analysis reveals a consistent lack of institutional preparedness across educational sectors, particularly in the areas of policy development, ethical oversight, and infrastructural support. In many contexts, students and educators engage with generative AI tools in the absence of clear guidelines, appropriate training, or meaningful safeguards. This governance vacuum intensifies risks around data misuse, academic dishonesty, and algorithmic opacity. Furthermore, there is evidence that many students do not perceive the use of AI-generated content as a breach of academic integrity, highlighting a critical disjunction between institutional expectations and learner perceptions.

To address these challenges, institutions must prioritise the development of transparent and context-sensitive policies that regulate AI use while safeguarding privacy and intellectual standards. The integration of ethical AI literacy into both student curricula and teacher professional development is essential to cultivate critical awareness, responsible usage, and digital fluency. Assessment practices should be redesigned to mitigate the risk of AI-facilitated misconduct, with greater emphasis on process-driven, authentic, and oral or collaborative forms of evaluation.

Equity must be central to the adoption of generative chatbots. This requires institutional investment in digital infrastructure to ensure access for all learners, alongside targeted support for marginalised groups. Developers and educators should collaborate to audit and mitigate algorithmic biases through inclusive design practices and regular scrutiny of training datasets. Data governance must be reimagined to comply with legal and ethical standards, ensuring privacy-by-design, transparency, and user agency in data use.

Finally, future research must extend beyond short-term case studies to include longitudinal and participatory inquiries that reflect the diversity of educational systems, particularly in underrepresented regions. A more global and inclusive research agenda is needed to ensure that the ethical discourse around AI in education remains relevant, equitable, and responsive to varied pedagogical, cultural, and institutional contexts.

The responsible integration of generative chatbots into teaching and learning hinges not only on technical competence but also on ethical clarity, institutional commitment, and shared pedagogical values. By aligning innovation with integrity, educational stakeholders can foster AI-enhanced learning environments that are just, inclusive, and resilient.

References

1. Al-kfairy, M., Mustafa, D., Kshetri, N., Insiew, M., & Alfandi, O. A Systematic Review and Analysis of Ethical Challenges of Generative Ai: An Interdisciplinary Perspective. *Available at SSRN 4833030*. <https://ssrn.com/abstract=4833030> or <http://dx.doi.org/10.2139/ssrn.4833030>
2. AL-Smadi, M. (2023). ChatGPT and Beyond: The Generative AI Revolution in Education. *arXiv*. <https://doi.org/10.48550/arXiv.2311.15198>
3. Bayly-Castaneda, K., Ramirez-Montoya, M. S., & Morita-Alexander, A. (2024, August). Crafting personalized learning paths with AI for lifelong learning: a systematic literature review. In *Frontiers in Education* (Vol. 9, p. 1424386). Frontiers. <https://doi.org/10.3389/educ.2024.1424386>
4. Boateng, O., & Boateng, B. (2025). Algorithmic bias in educational systems: Examining the impact of AI-driven decision making in modern education. *World Journal of Advanced Research and Reviews*, 25(1), 2012-2017. <https://doi.org/10.30574/wjarr.2025.25.1.0253>
5. Bukar UA, Sayeed MS, Razak SFA, Yogarayan S, Amodu OA. 2024. An integrative decision-making framework to guide policies on regulating ChatGPT usage. *PeerJ Computer Science* 10:e1845 <https://doi.org/10.7717/peerj-cs.1845>
6. Chan, C.K.Y., Hu, W. (2023). Students' voices on generative AI: perceptions, benefits, and challenges in higher education. *Int J Educ Technol High Educ* 20, 43. <https://doi.org/10.1186/s41239-023-00411-8>
7. Cotton, D. R. E., Cotton, P. A., & Shipway, J. R. (2023). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International*, 61(2), 228-239. <https://doi.org/10.1080/14703297.2023.2190148>
8. Davar, N. F., Dewan, M. A. A., & Zhang, X. (2025). AI Chatbots in Education: Challenges and Opportunities. *Information*, 16(3), 235. <https://doi.org/10.3390/info16030235>
9. Elkhatat, A.M. (2023). Evaluating the authenticity of ChatGPT responses: a study on text-matching capabilities. *Int J Educ Integr* 19, 15. <https://doi.org/10.1007/s40979-023-00137-0>

10. Evangelista, E. D. L. (2025). Ensuring academic integrity in the age of ChatGPT: Rethinking exam design, assessment strategies, and ethical AI policies in higher education. *Contemporary Educational Technology*, 17(1), ep559. <https://doi.org/10.30935/cedtech/15775>
11. Golda, A., Mekonen, K., Pandey, A., Singh, A., Hassija, V., Chamola, V., & Sikdar, B. (2024). Privacy and security concerns in generative AI: A comprehensive survey. *IEEE Access*, 12, 1-[Last page]. <https://doi.org/10.1109/ACCESS.2024.3381611>
12. Gruenhagen, J. H., Sinclair, P. M., Carroll, J.-A., Baker, P. R. A., Wilson, A., & Demant, D. (2024). The rapid rise of generative AI and its implications for academic integrity: Students' perceptions and use of chatbots for assistance with assessments. *Computers & Education: Artificial Intelligence*, 7, 100273. <https://doi.org/10.1016/j.caeai.2024.100273>
13. Halaweh, M. (2023). ChatGPT in education: Strategies for responsible implementation. *Contemporary educational technology*, 15(2).
14. Imran, M., & Almusharraf, N. (2023). Analyzing the role of ChatGPT as a writing assistant at higher education level: A systematic review of the literature. *Contemporary Educational Technology*, 15(4), ep464. <https://doi.org/10.30935/cedtech/13605>
15. Li, M., Enkhtur, A., Cheng, F., & Yamamoto, B. A. (2023). Ethical implications of ChatGPT in higher education: A scoping review. *arXiv preprint arXiv:2311.14378*. <https://doi.org/10.48550/arXiv.2311.14378>
16. Li, M., Enkhtur, A., Cheng, F., & Yamamoto, B. A. (2023). Ethical implications of ChatGPT in higher education: A scoping review. *arXiv*. <https://arxiv.org/abs/2311.14378>
17. Merino-Campos, C. (2025). The impact of artificial intelligence on personalised learning in higher education: A systematic review. *Trends in Higher Education*, 4(2), 17. <https://doi.org/10.3390/higheredu4020017>
18. OECD. (2024). *OECD Recommendation on the Ethics of Artificial Intelligence in Education*. OECD Publishing. <https://www.oecd.org/education/oecd-recommendation-on-the-ethics-of-ai-in-education.htm>
19. Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ... & Moher, D. (2021). Updating guidance for reporting systematic reviews: development of the PRISMA 2020 statement. *Journal of clinical epidemiology*, 134, 103-112. <https://doi.org/10.1136/bmj.n71>
20. Pitts, G., Marcus, V., & Motamedi, S. (2025). Student Perspectives on the Benefits and Risks of AI in Education. *arXiv preprint arXiv:2505.02198*. <https://doi.org/10.48550/arXiv.2505.02198>
21. UNESCO. (2023). *Guidelines for the governance of generative AI in education and research*. United Nations Educational, Scientific and Cultural Organization. <https://unesdoc.unesco.org/ark:/48223/pf0000386796>
22. Vartiainen, H., Valtonen, T., Kahila, J., & Tedre, M. (2025). ChatGPT and imaginaries of the future of education: insights of Finnish teacher educators. *Information and Learning Sciences*, 126(1/2), 75-90. <https://doi.org/10.1108/ILS-10-2023-0146>
23. Wang, H., Dang, A., Wu, Z., & Mac, S. (2023). Generative AI in Higher Education: Seeing ChatGPT through universities' policies, resources, and guidelines [Preprint]. *arXiv*. <https://doi.org/10.48550/arXiv.2312.05235>
24. Williams, R. T. (2024, January). The ethical implications of using generative chatbots in higher education. In *Frontiers in Education* (Vol. 8, p. 1331607). Frontiers Media SA. <https://doi.org/10.3389/feduc.2023.1331607>
25. Yan, L., Sha, L., Zhao, L., Li, Y., Martinez-Maldonado, R., Chen, G., & Han, Y. (2023). Practical and ethical challenges of large language models in education: A systematic scoping review. *arXiv*. <https://arxiv.org/abs/2303.13379>
26. Yan, L., Sha, L., Zhao, L., Li, Y., Martinez-Maldonado, R., Chen, G., ... & Gašević, D. (2024). Practical and ethical challenges of large language models in education: A systematic scoping review. *British Journal of Educational Technology*, 55(1), 90-112. <https://doi.org/10.48550/arXiv.2303.13379>
27. Zhai, C., Wibowo, S. & Li, L.D. (2024). The effects of over-reliance on AI dialogue systems on students' cognitive abilities: a systematic review. *Smart Learn. Environ.* **11**, 28. <https://doi.org/10.1186/s40561-024-00316-7>
28. Zhang, J., Song, W. & Liu, Y. (2025). Cognitive bias in generative AI influences religious education. *Sci Rep* **15**, 15720. <https://doi.org/10.1038/s41598-025-99121-6>

Figures

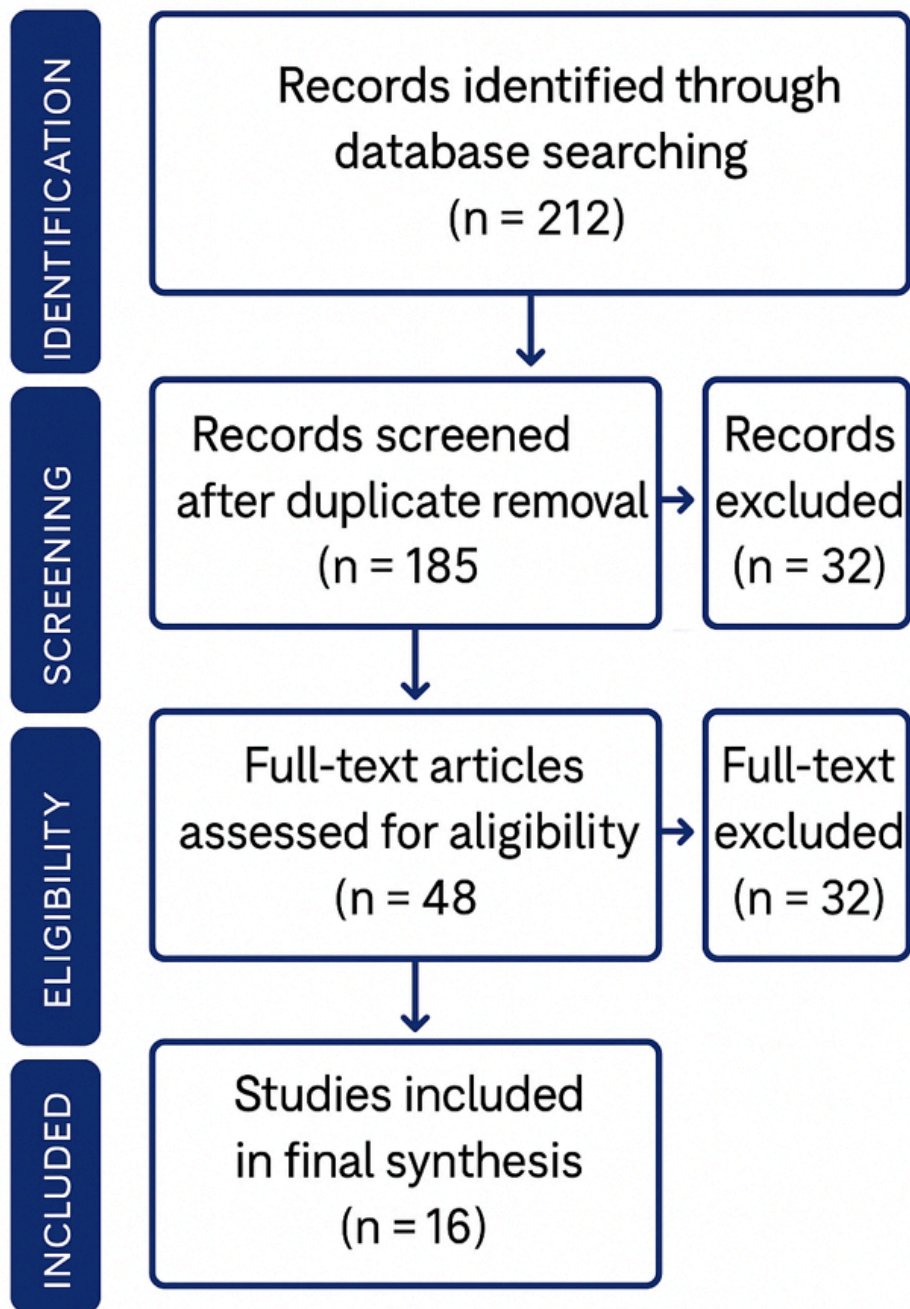


Figure 1

PRISMA Flow Diagram of Study Selection Process