

The Disclosure-Detection Nexus in Generative AI: An Integrated Analysis of Ethical Implications in Academia

Author(s): Jacob Holland

Botswana Accountancy College, Gaborone, Botswana

Corresponding Author: jacobh@bac.ac.bw

Received: December, 2025 Published: February, 2026

ARTICLE INFO

Keywords:

Artificial Intelligence; Disclosure; Detection; Ethical AI; GenAI

© 2026 by the Author(s). This open-access article is distributed under a Creative Commons Attribution (CC-BY) 4.0 license, making research freely available to the public and supporting a greater global exchange of knowledge and human experiments.



ABSTRACT

The rapid integration of Generative Artificial Intelligence (GenAI) into academic settings presents a complex ethical landscape, necessitating a critical examination of how institutions can effectively balance transparent AI usage (disclosure) with the need to identify misuse (detection). This study employs a comprehensive structured literature review utilizing thematic analysis to explore the multifaceted ethical implications of GenAI. Key issues synthesized include academic integrity, evolving notions of authorship, equity of access, intellectual property rights (IPRs), data privacy, and human agency, particularly highlighting the challenges posed by the "black box" nature of GenAI models, the potential for deskilling, and impacts on critical thinking and trust. This study underscores the inherent conflicts and interdependencies between disclosure and detection, noting the significant unreliability and biases of current AI detection tools that often lead to false positives and erode trust. It argues for a strategic shift towards a synergistic balance, where robust disclosure fosters transparency and pedagogical integrity, minimizing reliance on unreliable detection as a reactive measure. Actionable recommendations are provided for policymakers, educators, and technology developers to foster responsible AI integration, preserve academic values, and mitigate associated risks. The study concludes by outlining crucial future research directions, including longitudinal impact studies and empirical evaluations of assessment redesigns, to address the evolving challenges posed by GenAI in higher education.

1.0 Introduction

The advent of Generative Artificial Intelligence (GenAI) has marked a profound paradigm shift within academia, presenting both transformative opportunities and an increasingly complex ethical landscape (Symeou et al., 2025). Unlike prior AI-guided learning systems, which largely enhance interactive and personalized learning (Dingal et al., 2024; Kyambade et al., 2025), GenAI's capacity for autonomous content generation fundamentally challenges the bedrock principles of higher education. This rapid integration compels institutions to confront a critical dilemma: how to effectively balance the imperative for transparent AI usage (disclosure) with the need to identify and mitigate misuse (detection) (El Ali et al., 2024). This delicate equilibrium directly impacts core academic values, raising urgent ethical questions regarding academic integrity, evolving notions of authorship, equity of access, intellectual property rights (IPRs), and data privacy (Ardito, 2024; Bashir & Lapshun, 2025). The widespread availability of GenAI tools has not merely sparked discussion but has ignited a pressing, multifaceted debate globally within higher education globally (Ioku et al., 2024; Nguyen et al., 2022; Zhao et al., 2024) regarding the future of learning, assessment, and research ethics.

Despite the widespread acknowledgment of these ethical challenges, a critical void persists in the comprehensive analysis of the inherent conflicts and interdependencies between

disclosure and detection (Ardito, 2024; Benarab, 2024; Fontenot, 2025; Weaver, 2024). Current scholarly discourse frequently treats transparent AI usage and the identification of misuse as separate and isolated issues, thereby overlooking their dynamic and often conflicting relationships (Ali et al., 2024; Yu et al., 2023). Our integrated analysis specifically reveals that the perceived reliability of detection tools inadvertently stifles the development of robust disclosure policies, creating a vicious cycle of mistrust, or how the ethical implications across academic integrity, equity, and IP are mutually reinforced through the disclosure-detection dynamic. This fragmented perspective hinders the development of practical cohesive frameworks that institutions need to effectively manage. For instance, while some empirical research indicates positive student attitudes towards GenAI, it concurrently highlights significant ethical concerns and the urgent need for tailored AI literacy training (Vieira & Mesquita, 2025).

The inability to maintain these crucial balance risks undermines academic integrity, stifles innovation, and fosters environments of mistrust within educational communities. This fundamental tension is not isolated; it is significantly intertwined with and exacerbates other critical ethical concerns. For example, the imperative for transparent disclosure intersects directly with complex intellectual property (IP) issues surrounding AI-generated content (Wen, 2024; Xu, 2025), while the challenges of effective detection are inextricably linked to concerns about equitable access and algorithmic bias (Boateng & Boateng, 2025), which can disproportionately affect marginalized groups (Shahvaroughi Farahani & Ghasemi, 2024). Thus, a deeper understanding of these interconnections is essential to formulate comprehensive, fair, and sustainable solutions.

This study directly addresses this critical gap by providing a comprehensive and integrated analysis of the ethical ramifications of GenAI in academia, specifically focusing on the intricate and often conflicting balance between the disclosure and detection of AI-generated content. Beyond merely synthesizing existing concerns, this integrated approach uniquely demonstrates how the reciprocal failures of detection tools necessitate a proactive, trust-based disclosure framework that redefines institutional roles, or the specific systemic interdependencies between IP, equity, and pedagogical integrity that are overlooked in fragmented discussions. The study seeks to provide practical actionable recommendations for policymakers, educators, and technology developers, thereby contributing to the development of AI integration strategies that preserve academic values while mitigating risks.

Section 2 outlines the methodology for the literature review. Section 3 details the key ethical implications identified in the literature that inform Section 4's discussion of navigating the disclosure-detection nexus. Section 5 provides recommendations for various stakeholders. Section 6 presents the limitations of the study. Finally, Section 7 concludes and presents critical future research directions.

2.0 Methodology

This study employs a comprehensive literature review approach to critically synthesize the ethical implications of GenAI in academia, specifically focusing on the dynamic balance between the disclosure and detection of plagiarism. This methodology was designed to systematically identify, evaluate, and synthesize existing academic literature and policy documents to provide a robust understanding of the current landscape and provide actionable recommendations.

2.1. Literature Search Strategy

A literature search was conducted across multiple academic databases to ensure broad coverage of relevant publications. The primary databases utilized were EBSCOhost (including Academic Search Ultimate, ERIC, and PsycINFO) (n=60 records), Taylor and Francis Online (n=40 records), and institutional repositories (n=90 records). This yielded an initial total of 190 records. These databases were selected for their extensive coverage of

educational, social science, and technology-related literature, which is pertinent to the interdisciplinary nature of Gen-AI ethics in academia.

The search strategy involved a combination of keywords and Boolean operators to identify a wide range of relevant studies in the field. Key search terms included: "generative AI," "large language models," "LLMs," "academic integrity," "AI ethics," "educational technology," "disclosure," "detection," "plagiarism," "copyright," "intellectual property," "equity," "access," "privacy," "higher education," and "university." These terms were combined using "AND" and "OR" operators to construct comprehensive search strings (e.g., ("generative AI" OR "large language models" OR "LLMs") AND ("academic integrity" OR "AI ethics" OR "plagiarism") AND ("higher education" OR "university")).

To capture the progression of AI in education, the search was confined to publications from 2000 to 2025, with particular emphasis on recent advancements in GenAI owing to its rapid development over the past few years.

2.2. Inclusion and Exclusion Criteria

Specific inclusion and exclusion criteria were applied to ensure the relevance and quality of the synthesized literature.

2.2.1. Inclusion Criteria

1. Peer-reviewed journal articles, institutional reports, policy documents, and relevant conference proceedings were included.
2. Only English-language publications were included.
3. Studies explicitly address the ethical implications, challenges, opportunities, and policy responses related to Gen-AI in academic settings.
4. Empirical and theoretical studies have focused on the context of higher education contexts (Ardito, 2024; Bashir & Lapshun, 2025).
5. Documents discussing aspects of disclosure of AI use, detection of AI-generated content, academic integrity, equity of access, IP, data privacy, and Information and Communication Technology (ICT) infrastructure in relation to GenAI (Francis et al., 2025; Info & Potter, 2024).

2.2.2. Exclusion Criteria

1. Publications that were not peer-reviewed (e.g., blog posts, opinion pieces without academic rigor, and news articles) unless they were official institutional reports or policy documents.
2. Studies have primarily focused on AI applications outside education (e.g., medical diagnostics and industrial automation), without direct implications for academic contexts.
3. Publications that did not specifically address Generative AI (e.g., older AI-guided learning systems without generative capabilities) unless they provided foundational ethical considerations highly relevant to GenAI.
4. Duplicate records across databases.

2.3. Literature Screening and Selection

The initial search results were exported to a reference-management tool. After removing 70 duplicate records, a total of 120 unique records underwent a multi-stage screening process.

1. Title and Abstract Screening: Of the 120 unique records, the author screened titles and abstracts against the inclusion and exclusion criteria. This led to the exclusion of 20 records, primarily due to irrelevance to GenAI, focus outside higher education and

non-English language. 100 records were deemed potentially relevant and advanced to full-text review.

2. Full-Text Review: The full texts of the 100 potentially relevant articles were retrieved and independently reviewed. 20 articles were subsequently excluded after full-text assessment, with the primary reasons being a lack of ethical focus, an unsuitable methodology, and non-peer-reviewed status. 80 studies ultimately met all inclusion criteria and were included in the synthesis.

To enhance the reliability of the screening process, a second independent reviewer was engaged to screen a random sample of 10% of the titles and abstracts, and 20% of the full-text articles. Initial discrepancies were discussed and resolved by consensus, achieving Cohen's Kappa values of 0.78 for title/abstract screening and 0.82 for full-text review, demonstrating substantial inter-rater agreement.

2.4. Study Selection Flow

The rigorous, multi-stage process for identifying, screening, and selecting literature is summarized in Figure 1, presented in accordance with the PRISMA 2020 guidelines. This diagram illustrates the flow of information through the different phases of this systematic literature review, detailing the number of records identified, screened, and ultimately included in the synthesis, along with reasons for exclusion at each stage.

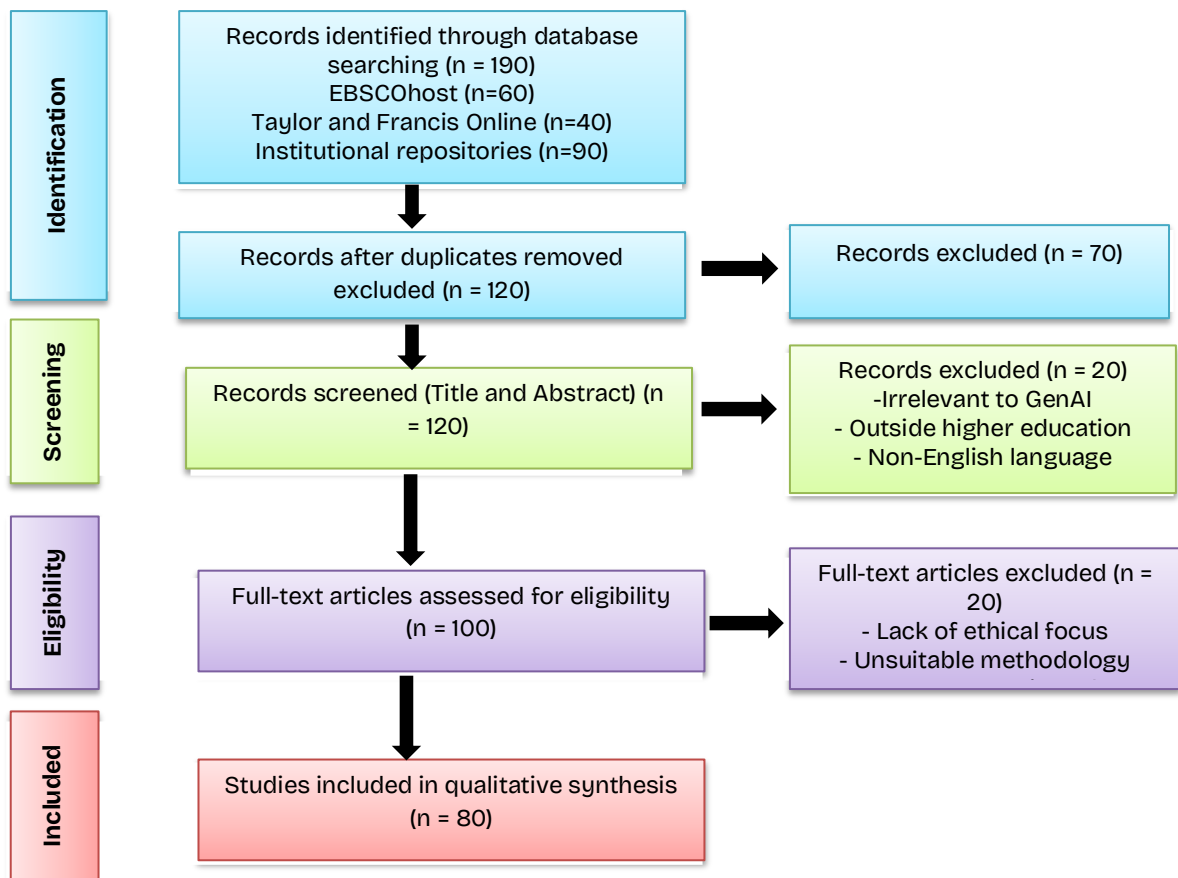


Fig 1: PRISMA 2020 Flow Diagram for Literature Selection

Source: Adapted from PRISMA 2020 Statement.

2.5. Data Extraction and Quality Assessment

For each included study, relevant data were systematically extracted using a predefined data-extraction form. The extracted information included the following.

1. Author(s), publication year, source.
2. Study design (e.g., empirical, theoretical, review, or case study).
3. Key ethical issues addressed (e.g., academic integrity, equity, IP, and privacy).
4. Specific findings related to GenAI disclosure and detection.
5. Identifying challenges and opportunities.
6. Recommendations for stakeholders (e.g., educators, policymakers, and students).

2.6. Analysis and Synthesis Approach

The extracted data were subjected to thematic analysis (Braun & Clarke, 2006) to identify recurrent patterns, key concepts, and overarching themes across a diverse body of literature. This involved:

1. Familiarization: The extracted data were thoroughly read and re-read.
2. Initial Coding: Generating initial codes from the data that captured distinct features or ideas related to Gen AI ethics, disclosure, and detection. For example, codes such as "AI detection tool limitations," "false positives," "student anxiety," "disclosure policy challenges," "equity of access," and IP ownership were created.
3. Searching for Themes: Grouping similar codes into broader, overarching themes. For example, codes related to "AI detection tool limitations," "false positives," and "bias against non-native speakers" coalesced into the theme of "Limitations of Detection Technologies."
4. Reviewing themes: The identified themes were refined and defined to ensure that they accurately reflected the data and addressed the research questions.
5. Defining and Naming Themes: Clear and concise names and definitions were developed for each theme, ensuring that they were distinct, yet interconnected.

The synthesis specifically focused on cross-disciplinary integration, drawing perspectives from education, management, and technological ethics to enable a holistic understanding of the ethical landscape of GenAI (Ardito, 2024; Bashir & Lapshun, 2025). Critical evaluation emphasizes identifying convergences and divergences in findings, highlighting areas of consensus, ongoing debate, and gaps in the literature. For example, recurring concerns about the accuracy and fairness of AI detection tools (Ardito, 2024; Bayraktar et al., 2025) were synthesized with pedagogical strategies advocating for robust assessment design and AI literacy (Al-Ali & Miles, 2025; Eleftheriou et al., 2025) to propose a balanced approach. This review critically examined how various studies addressed the central tension between encouraging transparent AI use (disclosure) and policing unauthorized applications (detection), providing examples from different contexts such as Saudi Arabian universities (Vieira & Mesquita, 2025) and built environment disciplines (Adebayo et al., 2025).

2.7. Limitations

Although a second independent reviewer was engaged for a substantial sample of articles during screening (10% of titles/abstracts, 20% of full texts), achieving significant inter-rater agreement (Cohen's Kappa values of 0.78 and 0.82), most of the literature screening, selection, data extraction, and qualitative assessment was primarily conducted by a single reviewer. This introduces an inherent degree of subjective interpretation, particularly for the unsampled portions, potentially limiting the breadth of perspectives in those cases.

To actively mitigate confirmation bias, a structured reflective journal was maintained throughout the review process. This involved documenting initial assumptions, critically appraising findings that challenged these assumptions, and regularly reviewing coding decisions against predefined criteria. This systematic approach to reflective practice aimed to enhance objectivity in interpretation and synthesis.

Given the broad and interdisciplinary scope of this study, encompassing diverse literature from empirical studies to theoretical frameworks and policy documents, a formal meta-analysis or quantitative quality assessment tool was not applied. However, to enhance transparency and reproducibility of the selection process, a PRISMA-style flow diagram (Figure 1) has been included. The analytical process involved a qualitative assessment of relevance, rigor, and contribution during the data extraction phase, prioritizing studies with clear methodologies, robust arguments, or significant empirical findings.

3.0 Results and Key Ethical Implications

The integration of GenAI into academic practices while concurrently offering transformative potential introduces a complex and evolving ethical dilemma. These challenges extend beyond mere technological adoption and touch on the core principles of academic integrity, fairness, human agency, and institutional trust. Understanding these multifaceted implications is paramount for developing robust governance frameworks and for ensuring responsible innovation.

3.1. Academic Integrity, Authorship and Misconduct

GenAI fundamentally challenges traditional definitions of academic integrity, authorship, and originality, thereby complicating the detection and prevention of misconduct (Ardito, 2024; Dingal et al., 2024). The ease with which GenAI can produce sophisticated content blurs the line between legitimate AI-assisted work and outright academic dishonesty.

3.1.1. Evolving Notions of Authorship and Originality

The pervasive capability of Generative AI (GenAI) to produce diverse content fundamentally challenges established academic conventions of authorship and originality, presenting significant practical hurdles for institutions. A primary challenge lies in the difficulty of attributing intellectual ownership, as the distinction between human and AI contributions blurs, complicating the assessment of individual learning (Bashir & Lapshun, 2025). This ambiguity directly affects academic integrity, requiring institutions to redefine acceptable AI integration into student work and research (Abubakar et al., 2024). For instance, practical challenges emerge in evaluating student submissions where AI has refined expression or generated initial drafts, making it difficult to discern students' genuine intellectual input from AI assistance. Emerging institutional responses, as noted in the literature, include developing new guidelines for acknowledging AI use and exploring alternative assessment methods that prioritize process and critical thinking over product generation, thereby clarifying the boundaries of legitimate AI-supported work.

3.1.2. Limitations and Biases of AI Detection Tools

Overreliance on AI detection tools for police GenAI use presents significant ethical and practical challenges. These tools frequently demonstrate vulnerabilities, including susceptibility to "paraphrasing attacks" and "parameter tweaking," which undermines their reliability (Perkins et al., 2024). Crucially, detection tools often struggle to differentiate between legitimate AI-assisted writing (e.g., refining expressions and brainstorming) and malpractice (e.g., generating entire assignments) (Barrot & Aranda, 2025; Info, 2024). These technical limitations often lead to high rates of false positives, disproportionately affecting diverse student populations and raising serious concerns about fairness (Elkhatat et al., 2023; Fisk, 2024). Empirical evidence, such as studies from Lebanese higher education, indicates that while

students acknowledge the use of AI, the efficacy of detection tools is widely questioned, underscoring a significant gap in effective policing (Vieira & Mesquita, 2025).

3.1.3. Impacts on Learning, Critical Thinking and Trust

The rapid integration of GenAI, along with the limitations of AI detection tools, poses a significant threat to both the quality of learning and foundational trust within academic institutions. The threat of false accusations due to flawed detection tools can induce significant anxiety and stress among students, fostering a climate of mistrust rather than academic honesty (Dalalah & Dalalah, 2023; Giray, 2024). This punitive approach, which focuses on surveillance, risks eroding the foundational trust between students and institutions, thereby diverting students' focus from genuine learning to detection anxieties. This concern is particularly prevalent in disciplines that require complex reasoning and synthesis.

Furthermore, the widespread accessibility of GenAI tools, while offering support, poses a substantial risk to the development of essential critical thinking, analytical, and problem-solving skills (Adebayo et al., 2025). If students become overly dependent on AI to generate ideas, structure arguments, or even perform basic calculations, their capacity for independent thinking and deep learning may diminish (Eleftheriou et al., 2025). Several studies further emphasize that overreliance on AI can diminish critical thinking, underscoring the need for ethical governance that prioritizes learning outcomes over simple detection (Lamberti et al., 2025; Pitts et al., 2025; Premkumar et al., 2024). Ultimately, these intertwined issues of anxiety, mistrust, and potential for deskilling highlight the critical need for ethical governance that prioritizes authentic learning outcomes and fosters a supportive academic environment.

3.2. Equity, Access, and Digital Divides

The integration of GenAI in academia risks exacerbating existing educational inequalities if access to these technologies and the necessary digital literacy skills are not equally distributed (Bittle & El-Gayar, 2025).

3.2.1. Disparities in Resource Availability

Access to advanced GenAI tools often correlates with socioeconomic status and geographical location. Students in resource-limited settings or those without reliable Internet access and powerful computing devices may face significant barriers to engaging in these technologies (James & Andrews, 2024). This creates a "digital divide" in AI adoption, where some students gain an unfair advantage in learning and academic performance while others are left behind (Qaribilla et al., 2024). For instance, findings from Lebanese universities revealed that only a minority of students perceive their institutions as having sufficient AI resources, highlighting infrastructural challenges (Vieira & Mesquita, 2025).

3.2.2. AI Literacy and Training Gaps

Equitable integration requires robust AI literacy beyond mere access to tools. Many students and educators lack adequate training in prompt engineering, understanding AI bias, and ethical AI use (Choi et al., 2025; Yeter et al., 2024). This gap in competency can prevent individuals from effectively leveraging the benefits of GenAI or, conversely, lead to unintentional misuse. Curricula must evolve to equip all students with the skills to critically engage with AI, ensuring that the benefits of AI-enhanced learning are broadly accessible, rather than limited to a privileged few (Bayraktar et al., 2025).

3.2.3. Algorithmic Bias and Fairness

GenAI models are trained on vast datasets that often reflect existing societal biases. This can lead to AI systems that perpetuate or even amplify stereotypes, producing outputs that are unfair, discriminatory, or culturally insensitive (Chaudhuri & Petkovic, 2025; Chugh et al., 2025). Such biases can manifest in content generation, personalized learning recommendations, and even the detection of AI-generated text, further disadvantaging marginalized groups. Ethical frameworks must explicitly address the identification and mitigation of algorithmic bias to ensure fairness in all academic AI applications.

3.3. Intellectual Property, Data Privacy, and Security

The extensive data-processing capabilities of GenAI and its capacity to generate content introduce complex ethical and legal challenges related to IP, privacy, and cybersecurity.

3.3.1. Ownership of AI-Generated Content

The legal frameworks surrounding IP were not designed for a world in which machines could generate creative work (Mazzi, 2024; Picht & Thouvenin, 2023). Ambiguities persist regarding who owns the copyright to AI-generated academic content: the user, AI developer, or neither. This uncertainty affects researchers, students, and institutions, particularly in fields that require original contributions and patentable innovation. Clear institutional policies are needed to define ownership and usage rights for AI-assisted outputs (Mahrishi et al., 2025).

3.3.2. Cybersecurity Vulnerabilities

Increasing reliance on AI platforms has introduced new cybersecurity risks. Malicious actors can exploit vulnerabilities in AI systems to compromise data, introduce misinformation, or disrupt academic operations (Mbah & Evelyn, 2024). Institutions must invest in advanced cybersecurity measures and continuous monitoring to safeguard their digital infrastructure and integrity of AI-enhanced educational activities (Kyambade et al., 2025).

3.3.3. Transparency, Intellectual Property, and Data Privacy Risks

The inherent "black box" nature of many advanced GenAI models, where the internal decision-making processes and the sources from which they derive their outputs are opaque and difficult to interpret (Balasubramaniam et al., 2023), significantly exacerbates IP and data privacy risks. Regarding IP, this lack of transparency makes it challenging to ascertain the origins of the AI training data. This opacity creates uncertainty about whether the copyrighted material was used without proper licensing or attribution during the model development. Consequently, any content generated by such an AI could inadvertently infringe on existing copyrights, leading to a potential legal liability for users and institutions (Buick, 2024). Furthermore, when students submit their original, unpublished work, such as thesis drafts, into proprietary 'black box' AI tools, there is a significant, often uncommunicated, risk that the terms of service may grant the AI developer broad licenses to use, reproduce, or adapt the student's input data, effectively compromising or even ceding the student's IPRs to their own academic output. For example, if an AI generates a research paper or creative work that closely resembles a copyrighted piece from its nontransparent training data, it becomes difficult to defend against claims of infringement because the exact influence of the source material cannot be traced or disproven. Furthermore, the "black box" complicates the determination of authorship and ownership (as discussed above); without insight into how the AI processed prompts and synthesized information, it is harder to differentiate between genuine human intellectual contributions and AI-driven reproduction, thereby muddying IP claims for AI-assisted works (Mazzi, 2024; Picht & Thouvenin, 2023).

Regarding data privacy, the nontransparent operations of GenAI models pose substantial threats to confidentiality and data protection. When users input sensitive personal information, proprietary research data, or confidential institutional records into these models, the lack of visibility in the AI's internal processing means that users and institutions cannot verify how these data are handled, stored, or retained within the model's parameters. This opacity makes it difficult to ensure compliance with stringent data protection regulations, such as the General Data Protection Regulation (GDPR), as auditing the model's data processing mechanisms for privacy safeguards (e.g., anonymization and deletion of personal data) becomes practically impossible (Granata et al., 2024).

There is a significant risk that the model might “memorize” and inadvertently reproduce private data from its training set or user inputs, leading to unintended data exposure or a breach of confidentiality (Gupta & Nyamapfene, 2025; Jose, 2024). Without explainability, it is impossible to confirm that personal data are not being repurposed or inadvertently exposed, creating persistent privacy vulnerability. For university administrators, this presents a significant management tension: how to encourage the innovative use of AI tools for learning and research, while simultaneously safeguarding students' IP, particularly for high stakes work like theses or dissertations. This tension is heightened by the often complex and obscure terms of service agreements of proprietary AI tools, which students may accept without fully understanding the implications for their IP. Users must understand the limitations and potential biases of AI outputs to critically evaluate them rather than accepting them uncritically (Chaudhry et al., 2022).

3.4. Human Agency, Autonomy, and Well-being

The ethical implications of GenAI extend to its impact on human agency, individual autonomy, and the overall well-being of academic stakeholders.

3.4.1. Deskillling and Over-reliance

As GenAI tools become more sophisticated, there is a risk that individuals may become over-reliant on them for tasks that traditionally require human skills and effort. This “deskilling” could diminish human capabilities in areas such as writing, critical analysis, and problem-solving (Eleftheriou et al., 2025; Sousa & Cardoso, 2025). Maintaining a balance in which AI augments rather than replacing human capabilities is crucial for fostering continued intellectual growth.

3.4.2. Emotional and Psychological Impact

Beyond the anxiety of false accusations (Ardito, 2024), the pervasive presence of GenAI can affect student and faculty well-being in other ways. The constant pressure to perform against AI-generated benchmarks or the feeling of being replaced by technology can contribute to stress and burnout. Conversely, the successful integration of AI can reduce mundane tasks, freeing time for more creative and impactful work if managed ethically (Al-Ali & Miles, 2025).

3.4.3. Pedagogical Shifts and Teacher Autonomy

Educators face the challenge of adapting their pedagogical strategies to ethically incorporate Gen-AI (Al-Ali & Miles, 2025; Bayraktar et al., 2025). This requires significant professional development and institutional support to ensure that teachers feel confident and competent in guiding students' ethical AI use rather than feeling threatened or overwhelmed. Preserving teacher autonomy in curriculum design and assessment remains crucial when navigating these technological shifts.

3.4.4. Accountability and Responsibility

Establishing clear lines of accountability and responsibility is complex in instances of AI errors or misuse. When AI generates factually incorrect information or biased content, who is responsible for academic or ethical breaches? This question is particularly salient in research, publications, and in critical decision-making processes. Ethical frameworks must delineate the responsibilities of users, developers, and institutions (Bashir & Lapshun, 2025).

4.0 Discussion: Navigating the Disclosure-Detection Nexus for Responsible GenAI Integration

Building on the key ethical implications identified in the literature (Section 3), this discussion synthesizes the existing research to explore strategies for successfully incorporating GenAI into academic environments. It focuses on the crucial interplay between transparent disclosure and effective detection, arguing for a synergistic approach in which strong disclosure practices can mitigate the challenges of unreliable detection and foster a more trustworthy and pedagogically sound academic ecosystem.

4.1. Fostering Transparent Disclosure: Cultivating Trust and Academic Partnership

Institutions are increasingly implementing policies (Wang et al., 2024) that require students and researchers to disclose the use of AI in academia. These policies range from voluntary acknowledgments to compulsory and explicit reporting within assignments or research outputs (Overono & Ditta, 2024; Resnik & Hosseini, 2025). Effective disclosure fosters transparency, encourages ethical AI use during the learning process, and supports academic authenticity. However, policies must be crafted carefully to avoid imposing undue burdens or raising privacy concerns regarding the specifics of AI interactions.

Disclosure policies require clear guidance and practical examples to be effective. This guidance should delineate the acceptable and unacceptable uses of AI, provide templates for citation and acknowledgment, and clarify the expectations for human oversight and intellectual contribution when AI tools are employed (Ugwu et al., 2024).

Pedagogical strategies such as the cognitive apprenticeship model (Al-Ali & Miles, 2025), can be instrumental in cultivating a culture of disclosure. By modelling responsible AI integration, coaching students through ethical dilemmas related to AI use, and fostering reflective practices, educators can empower students to transparently understand why and how to disclose AI assistance. For instance, an instructor might model how to cite AI use in a research proposal, and then coach students on proper disclosure in their own work, reinforcing this through peer review and feedback (Eleftheriou et al., 2025).

When students are genuinely empowered and educated on why and how to disclose their AI use and when academic cultures value transparency, the default institutional reaction can shift from suspicion and surveillance to trust and academic partnerships. Proactive disclosure, such as requiring an 'AI Use Log' detailing prompts, AI outputs, and subsequent human modifications, provides a transparent audit trail. This approach not only fosters honesty but also provides educators with valuable insights into students' learning processes with AI, moving beyond a simple 'yes/no' AI detection and significantly reducing the burden on unreliable detection methods.

4.2. Re-evaluating Detection: Shifting from Surveillance to Pedagogical Integrity

While disclosure is crucial for transparency, institutions also grapple with the need for reliable mechanisms to identify unauthorized AI-generated content (Sharma & Sharma, 2024). However, as discussed in Section 3, current AI detection tools face significant limitations, including susceptibility to "paraphrasing attacks" and "parameter tweaking," leading to high rates of false positives and disproportionately affecting diverse student populations (Perkins, Roe, et al., 2024). Overreliance on imperfect tools risks fostering a

climate of mistrust and anxiety, diverting focus from genuine learning (Dalalah & Dalalah, 2023; Giray et al., 2024; Wardat, 2024).

The inherent unreliability of automated detection necessitates a strategic shift to this approach. Instead of investing heavily in unreliable detection software, resources should be redirected towards educating students and faculty on ethical AI use, fostering AI literacy, and redesigning assessments that are less susceptible to AI generation. For example, implementing authentic assessments such as oral presentations, project-based learning, or critical reflection assignments, which demand human interpretation, creativity, and unique insights, can circumvent the limitations of GenAI (Chen & Lv, 2025). This approach not only indirectly addresses the detection challenge but also enhances educational experience by promoting higher-order thinking.

However, the implementation of this pedagogical shift presents challenges. Redesigning curricula and assessment strategies across an institution requires significant faculty development, time, and resources, which is a potential trade-off against the perceived ease of deploying an AI detection tool. Moreover, while disclosure policies aim to cultivate trust, there will inevitably be instances in which students will choose not to disclose. In such cases, a clear institutional response is required to uphold academic integrity without resorting to unreliable technological solutions. This may involve developing human-led review processes. These processes would involve trained academic staff systematically evaluating submissions for specific indicators such as abrupt shifts in writing style, inconsistent vocabulary or tone, lack of critical engagement with source material, or a disconnect between the complexity of the task and the depth of reasoning presented. This evaluation would move beyond surface-level text analysis to assess the underlying intellectual contribution. For instance, if a student submits an essay with a highly sophisticated introduction but then transitions to simplistic arguments and grammatical errors, a human reviewer might flag this inconsistency. Further investigation could involve comparing it to the student's previous work, or inviting the student to an interview to explain their thought process and specific sections of the essay, rather than relying on an AI detector's score.

4.3. The Synergistic Balance: Disclosure as a Proactive, Detection as a Responsive Strategy

Achieving a true "balance" between disclosure and detection is not about equal emphasis but rather about creating a complementary ecosystem where robust disclosure practices reduce the necessity for stringent, often unreliable, detection (Dalalah & Dalalah, 2023). Unlike other hybrid models that may equally weigh both components, our Synergistic Balance uniquely prioritizes pedagogical integrity and student agency as the foundational drivers for disclosure, rendering detection a secondary, human-augmented failsafe. Disclosure serves as a proactive preventative measure built on trust and education (Spirgi, 2025). This empowers students and researchers to transparently engage with GenAI, fostering a culture of academic honesty. Conversely, detection becomes a responsive, carefully considered strategy reserved for instances where disclosure is absent or academic integrity is genuinely questioned (Halaweh & Refae, 2024). This distinct prioritization, where proactive trust-building reduces the reliance on reactive policing, sets our 'Synergistic Balance' apart from approaches that maintain a more equal emphasis or sequential application of disclosure and detection.

This balance is dynamic: as AI capabilities evolve, policies and pedagogical approaches must adapt to them. Initially, a strong focus on detection might have seemed necessary, but as AI detection tools have proven unreliable, the balance has shifted towards empowering students through disclosure and robust pedagogy (Elkhatat et al., 2023). For instance, in a course emphasizing creative writing, the balance might heavily lean on the transparent disclosure of AI use for brainstorming, coupled with authentic assessments, such as live presentations or defense of work. Conversely, in foundational knowledge assessments

where independent recall is paramount, the balance might still necessitate some form of human-augmented detection alongside clear “no-AI” disclosure policies. Ultimately, integrating Gen-AI requires institutions to prioritize educational empowerment and integrity over a punitive, technology-driven approach.

To effectively implement the synergistic balance between disclosure and detection, and to prioritize educational empowerment and integrity, as advocated in the preceding discussion, a multi-pronged strategy is essential, requiring specific actionable responsibilities from various stakeholders.

5.0 Actionable Recommendations for Stakeholders

Achieving this balance involves a multipronged strategy with specific responsibilities for various stakeholders.

5.1. For Policymakers and Academic Institutions

1. **Develop Clear, Adaptive, and Inclusive Regulatory Frameworks:** Explicitly address IP ownership for AI-generated content, delineate legal liability for AI errors or biases, and establish robust data privacy and security protocols specifically tailored to GenAI's capabilities. Crucially, university administrators must proactively manage the risk of students inadvertently losing IPRs to their thesis work by running it through proprietary 'black box' AI tools. This entails reviewing the terms of service of commonly used AI platforms, advising students on the IP implications of using such tools for sensitive academic work, and potentially providing institutionally vetted or open-source AI alternatives where IPRs are clearly protected. These policies must balance innovation with protecting students' academic values and individual rights.
2. **Ensure Equitable Access and AI Literacy Training:** Promote equitable access to resources and comprehensive AI literacy training across all demographics to mitigate the digital divide.
3. **Supportive Policies for Disclosure:** Frame disclosure policies not as a confession of wrongdoing but as an integral part of academic practice, like citing sources. This encourages students to leverage AI ethically as a learning aid, while maintaining accountability for their intellectual contributions.
4. **Contextual and human-centered detection:** When concerns about AI misuse arise, investigations should rely on holistic evidence. This includes detailed writing-style analysis (e.g., comparing the submission to known student work or assessing for sudden shifts in complexity or tone), deep content understanding (e.g., evaluating the originality of ideas, depth of critical analysis, and proper citation practices), and structured student interviews (e.g., asking students to explain their reasoning, development process, or specific passages). This human-centered approach acknowledges the limitations of technology and prioritizes fairness in its application. For example, in a suspected case of AI misuse, an academic integrity officer might first review the student's past assignments for writing style consistency. If discrepancies are noted, they would then assess the submitted content for depth of critical thought and originality, cross-referencing sources. Finally, a structured interview with the student would allow them to explain their work process, demonstrate understanding of the subject matter, and clarify any suspicious elements, providing a comprehensive, human-centered assessment.

5.2. For Educators

1. **Prioritize Pedagogical Innovation through Robust Assessment Redesign:** Shift away from easily AI-generatable assignments towards authentic, higher-order thinking tasks (e.g., oral examinations, real-world case studies, process-based assignments) that inherently require human creativity and critical analysis.

2. **Implement Comprehensive Professional Development:** Equip educators with advanced AI literacy, prompt engineering skills, and strategies for ethically integrating AI into curricula while fostering student agency and critical engagement (e.g., through models such as TPTP).
3. **Clear Communication and Education:** Clearly articulate acceptable and unacceptable uses of GenAI, providing comprehensive training for students and faculty on AI literacy, and prompt engineering and ethical guidelines. This educational component should precede all the punitive measures.

5.3. For Technology Developers

1. **Commit to Ethical AI Development:** Develop AI tools that prioritize transparency, explainability, fairness, and interoperability.
2. **Implement Robust Bias Mitigation:** Integrate mechanisms for bias detection and mitigation, clearly communicate model limitations, and ensure secure data-handling practices.
3. **Engage in Participatory Design:** Collaborate with academic stakeholders to ensure that tools align with educational values and support ethical academic practices.

6.0. Study Limitations

Although this study offers a comprehensive synthesis of the ethical implications of GenAI in academia, it is important to acknowledge several inherent limitations that shape its findings and recommendations for future research.

1. **Dynamic and Rapidly Evolving Landscape:** GenAI is developing at an unprecedented rate. The ethical considerations, technological capabilities, and institutional responses discussed in this review represent a snapshot of the current discourse, primarily drawn from literature up to 2025. This rapid evolution means that some findings and recommendations may quickly become outdated as new tools emerge, policies are implemented, and research progress. For example, the efficacy and bias of AI detection tools, a central theme, are subject to continuous change and improvement, potentially altering the disclosure detection nexus over time.
2. **Reliance on Existing Literature:** As structured literature review, this study's insights are inherently dependent on the scope, quality, and potential biases of the included studies. Although efforts have been made to ensure comprehensive coverage, this review did not generate new empirical data or primary observations. Consequently, the findings reflect a synthesis of reported perspectives and evidence rather than a direct empirical investigation of the practical impacts or effectiveness of the proposed strategies within specific academic contexts.
3. **Breadth Over Depth in Technical Aspects:** This review focuses on the ethical implications and disclosure-detection nexus from a pedagogical and policy perspective. While touching upon the "black box" nature of GenAI models, it does not delve into the granular technical details of AI model architecture, training data specifics, or algorithmic mechanisms of detection tools. A deeper technical analysis, although beyond the scope of this review, could offer additional insights into the root causes of certain ethical challenges such as algorithmic bias or data privacy vulnerabilities.
4. **Generalizability Across Diverse Contexts:** While review aims for global relevance, the synthesis of the literature may be influenced by the predominant regions, types of institutions, or academic disciplines represented in existing research. The practical applicability of the recommendations may vary significantly across institutional

cultures, regulatory environments, and resource availabilities. For instance, recommendations for assessment redesign might face different implementation challenges in a large public university than in a specialized private college.

5. **Subjectivity in Thematic Interpretation:** Although rigorous thematic analysis approach was employed, the process of coding and theme development inherently involved the degree of subjective interpretation by the reviewer. Although efforts have been made to maintain objectivity and consistency, alternative interpretations of the same body of literature are possible.

7.0. Conclusion and Future Works

The rapid integration of GenAI into academia presents a transformative, yet ethically complex landscape, necessitating a delicate balance between fostering transparent AI use (disclosure) and developing nuanced strategies to detect misuse. This critical synthesis illuminates multifaceted ethical dilemmas, fundamentally challenging traditional notions of academic integrity by redefining authorship and exposing the unreliability and biases of current AI detection tools. It also highlighted how GenAI exacerbates existing educational inequalities through disparities in access and algorithmic biases while introducing significant legal ambiguities concerning IP and raising critical data privacy and cybersecurity risks. Furthermore, the “black box” nature of many GenAI models raises concerns regarding transparency and explainability, risking deskilling students, and diminishing critical thinking. Addressing these challenges requires interdisciplinary approaches that integrate legal, pedagogical, and technological perspectives and move beyond punitive detection towards proactive ethical engagement. Ultimately, by embracing these proactive and human-centered strategies, higher education can not only navigate the challenges of Gen-AI but also harness its transformative potential to foster a more equitable, innovative, and ethically robust academic environment.

Future Research Directions

Building on the insights of this synthesis, future research is vital to deepen our understanding and refine strategies for ethical Gen-AI integration.

1. **Longitudinal Impact Studies:** Conduct long-term empirical studies across diverse academic settings and disciplines (e.g., built environment, management education) to assess the longitudinal impacts of Gen-AI on learning outcomes, student cognitive development (e.g., critical thinking, problem-solving), academic culture, and equity, thereby providing essential evidence to inform adaptable, foresightful policy-making, and pedagogical innovations that truly support student growth.
2. **Effectiveness of Assessment Redesign:** Empirically evaluate the efficacy of various redesigned authentic assessment methods in mitigating Gen-AI misuse while simultaneously fostering higher-order thinking skills and student engagement. This should include cross-cultural comparisons to understand contextual nuances and provide evidence-based models for widespread adoption that uphold academic rigour and promote genuine learning.
3. **Psychological and Trust Impacts:** Investigate the long-term psychological effects of AI detection tools and mandatory disclosure policies on student well-being, anxiety levels, and foundational trust between students and academic institutions, which is vital for cultivating a supportive and ethical learning environment in which trust rather than suspicion is paramount.
4. **Ethical AI Literacy Frameworks:** Develop and evaluate comprehensive AI literacy frameworks tailored for specific academic contexts, focusing on pedagogical approaches that empower students and educators to critically engage with AI, understand its biases, and utilize it ethically and responsibly, thereby ensuring that

all academic stakeholders are equipped to navigate and contribute to a responsible AI integrated future.

Regulatory Evolution and IP: Further research is needed on evolving legal and regulatory frameworks, particularly concerning IP ownership of AI-generated academic content and the allocation of responsibility for AI-induced errors or biases in research. This should include comparative studies of international approaches to AI governance in higher education, to establish clear guidelines that protect creators, foster innovation, and ensure equitable accountability in the academic landscape.

References

- [1] Abubakar, U., Falade, A. A., & Ibrahim, H. A. (2024). Redefining student assessment in Nigerian tertiary institutions: The impact of AI technologies on academic performance and developing countermeasures. *Advances in Mobile Learning Educational Research*, 4(2), 1149–1159. <https://doi.org/10.25082/amler.2024.02.009>
- [2] Adebayo, A., Oyedokun, T., Oladiran, O., Hossain, M., & Jagun, Z. (2025). Analysing the built environment academics' perceptions of generative AI technology on teaching and learning practice. *Cogent Education*, 12(1). <https://doi.org/10.1080/2331186x.2025.2511034>
- [3] Al-Ali, S., & Miles, R. (2025). Upskilling teachers to use generative artificial intelligence: The TPTP approach for sustainable teacher support and development. *Australasian Journal of Educational Technology*. <https://doi.org/10.14742/ajet.9652>
- [4] Ali, I., Warraich, N. F., & Butt, K. (2024). Acceptance and use of artificial intelligence and AI-based applications in education: A meta-analysis and future direction. *Information Development*, 41(3), 859–874. <https://doi.org/10.1177/02666669241257206>
- [5] Ardito, C. G. (2024). Generative AI detection in higher education assessments. *New Directions for Teaching and Learning*, 2025(182), 11–28. <https://doi.org/10.1002/tl.20624>
- [6] Balasubramaniam, N., Kauppinen, M., Hiekkänen, K., Kujala, S., & Rannisto, A. (2023). Transparency and explainability of AI systems: From ethical guidelines to requirements. *Information and Software Technology*, 159, 107197. <https://doi.org/10.1016/j.infsof.2023.107197>
- [7] Barrot, J. S., & Aranda, M. R. R. (2025). Efficacy of AI-Text Detection Tools in Distinguishing Student-Produced, AI-Edited, and AI-Generated Essays. *Technology, Knowledge and Learning*. <https://doi.org/10.1007/s10758-025-09884-0>
- [8] Bashir, S., & Lapshun, A. L. (2025). Generative artificial intelligence integration in management education: application and ethical challenges. *Cogent Education*, 12(1). <https://doi.org/10.1080/2331186x.2025.2526436>
- [9] Bayraktar, B., Henry, D., & Taggart, J. (2025). Navigating the AI-enabled education landscape: A multifaceted approach to providing effective professional learning and support for educators. *Theory Into Practice*, 64(4), 421–433. <https://doi.org/10.1080/00405841.2025.2528543>
- [10] Benarab, I. H. (2024). Detection of AI-generated Writing in Students' Assignments: A Comparative Analysis of Some Tools' Reliability. *ATRAS Journal*, 5(3), 271–286. <https://doi.org/10.70091/atras/ai.17>
- [11] Bittle, K., & El-Gayar, O. (2025). Generative AI and Academic Integrity in Higher Education: A Systematic Review and Research Agenda. *Information*, 16(4), 296. <https://doi.org/10.3390/info16040296>
- [12] 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>
- [13] Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- [14] Buick, A. (2024). Copyright and AI training data—transparency to the rescue? *Journal of Intellectual Property Law and Practice*, 20(3), 182–192. <https://doi.org/10.1093/jiplp/jpae102>

- [15] 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>
- [16] Chaudhuri, A., & Petkovic, D. (2025). Issues and Concerns About GenAI and How We Can Smartly Use It in Education. *Proceedings of the AAAI Symposium Series*, 5(1), 284–285. <https://doi.org/10.1609/aaais.v5i1.35600>
- [17] Chen, Y., & Lv, C. (2025). The quality evaluation index system for AI-generated digital educational resources. *Frontiers in Humanities and Social Sciences*, 5(6), 49–57. <https://doi.org/10.54691/crzb0f70>
- [18] Choi, W. C., Ng, S. I., Choi, I. C., Lam, L. C., Chang, C. I., & Leong, K. I. (2025). Artificial Intelligence (AI) Literacy in Education: Definition, Competencies, Opportunities and Challenges. *Mdpi Ag*. <https://doi.org/10.20944/preprints202508.0497.v1>
- [19] Chugh, R., Kaisar, S., Sabrina, F., Turnbull, D., Azad, S., Morshed, A., Md Mamunur, R., & Subramani, S. (2025). The Promise and Pitfalls: A Literature Review of Generative Artificial Intelligence as a Learning Assistant in ICT Education. *Computer Applications in Engineering Education*, 33(2). <https://doi.org/10.1002/cae.70002>
- [20] Dalalah, D., & Dalalah, O. M. A. (2023). The false positives and false negatives of generative AI detection tools in education and academic research: The case of ChatGPT. *The International Journal of Management Education*, 21(2), 100822. <https://doi.org/10.1016/j.ijme.2023.100822>
- [21] Dingal, S. M. L., Pelandas, A. M. O., Cabrera, C. A., Wahing, M. R. A., Buenavista, P. E. M., Olalo, A. M., Balbin, C. J. P., Clamares, K. J. M., Pataganao, L. A. A., Buenafe, S. G. L., & Lorica, V. M. T. (2024). Artificial Intelligence (AI) Usage and Its Influence to the Students' Academic Writing: A Quantitative – Correlation Investigation. *International Journal of Research and Innovation in Social Science*, 8(4), 1621–1627. <https://doi.org/10.47772/ijriss.2024.804216>
- [22] El Ali, A., Naudts, L., Helberger, N., Venkatraj, K. P., Cesar, P., & Morosoli, S. (2024). Transparent AI Disclosure Obligations: Who, What, When, Where, Why, How. 43, 1–11. <https://doi.org/10.1145/3613905.3650750>
- [23] Eleftheriou, M., Fredrick, D., & Ahmer, M. (2025). Balancing ethics and support: Peer tutors' experiences with AI tools in student writing. *Contemporary Educational Technology*, 17(3), ep587. <https://doi.org/10.30935/cedtech/16554>
- [24] Elkhatat, A. M., Elsaid, K., & Almeer, S. (2023). Evaluating the efficacy of AI content detection tools in differentiating between human and AI-generated text. *International Journal for Educational Integrity*, 19(1). <https://doi.org/10.1007/s40979-023-00140-5>
- [25] Fisk, G. D. (2024). AI or Human? Finding and Responding to Artificial Intelligence in Student Work. *Teaching of Psychology*, 52(3), 314–318. <https://doi.org/10.1177/00986283241251855>
- [26] Fontenot, J. (2025). Artificial Intelligence Disclosure in Academic Nursing: A Framework for Editorial Policy and Practice. *Nurse Author & Editor*, 35(1). <https://doi.org/10.1111/nae.2.70002>
- [27] Francis, N. J., Jones, S., & Smith, D. P. (2025). Generative AI in Higher Education: Balancing Innovation and Integrity. *British Journal of Biomedical Science*, 81. <https://doi.org/10.3389/bjbs.2024.14048>
- [28] Giray, L. (2024). The Problem with False Positives: AI Detection Unfairly Accuses Scholars of AI Plagiarism. *The Serials Librarian*, 85(5–6), 181–189. <https://doi.org/10.1080/0361526x.2024.2433256>
- [29] Giray, L., Jacob, J., & Gumalin, D. L. (2024). Strengths, Weaknesses, Opportunities, and Threats of Using ChatGPT in Scientific Research. *International Journal of Technology in Education*, 7(1), 40–58. <https://doi.org/10.46328/ijte.618>
- [30] Granata, D., Mastroianni, M., Cantiello, P., Rak, M., & Salzillo, G. (2024). GDPR compliance through standard security controls: An automated approach. *Journal of High Speed Networks*, 30(2), 147–174. <https://doi.org/10.3233/jhs-230080>
- [31] Gupta, V., & Nyamapfene, A. (2025). Generative AI in Universities: Practices at UCL and Other Institutions, and the Path Forward. *Internet Reference Services Quarterly*, 29(1), 131–151. <https://doi.org/10.1080/10875301.2025.2453461>
- [32] Halaweh, M., & Refae, G. E. (2024). Examining the Accuracy of AI Detection Software Tools in Education. 186–190. <https://doi.org/10.1109/idsta62194.2024.10747004>

- [33] Info, A. (2024). Reviewing the performance of AI detection tools in differentiating between AI-generated and human-written texts: A literature and integrative hybrid review. *Journal of Applied Learning & Teaching*, 7(1). <https://doi.org/10.37074/jalt.2024.7.1.14>
- [34] Info, A., & Potter, M.-A. (2024). Generative Artificial Intelligence in distance education: Transformations, challenges, and impact on academic integrity and student voice. *Journal of Applied Learning & Teaching*, 7(1). <https://doi.org/10.37074/jalt.2024.7.1.41>
- [35] Ioku, T., Kondo, S., & Watanabe, Y. (2024). Acceptance of generative AI in higher education: A latent profile analysis of policy guidelines. Springer Science Business Media Llc. <https://doi.org/10.21203/rs.3.rs-4515787/v1>
- [36] James, T., & Andrews, G. (2024). Levelling the playing field through GenAI: Harnessing artificial intelligence to bridge educational gaps for equity and disadvantaged students. *Widening Participation and Lifelong Learning*, 26(3), 250–260. <https://doi.org/10.5456/wpll.26.3.250>
- [37] 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>
- [38] Kyambade, M., Namatovu, A., & Male Ssentumbwe, A. (2025). Exploring the evolution of artificial intelligence in education: from AI-guided learning to learner-personalized paradigms. *Cogent Education*, 12(1). <https://doi.org/10.1080/2331186x.2025.2505297>
- [39] Lamberti, W., Lawrence, S., White, D., Kim, S., & Abdullah, S. (2025). Pilot Study on Generative AI and Critical Thinking in Higher Education Classrooms. <https://doi.org/10.48550/arxiv.2509.00167>
- [40] Mahrishi, M., Siddiqui, M. K., & Abbas, A. (2025). Global Initiatives Towards Regulatory Frameworks for Artificial Intelligence (AI) in Higher Education. *Digital Government: Research and Practice*, 6(2), 1–9. <https://doi.org/10.1145/3672462>
- [41] Mazzi, F. (2024). Authorship in artificial intelligence-generated works: Exploring originality in text prompts and artificial intelligence outputs through philosophical foundations of copyright and collage protection. *The Journal of World Intellectual Property*, 27(3), 410–427. <https://doi.org/10.1111/jwip.12310>
- [42] Mbah, G., & Evelyn, A. (2024). AI-powered cybersecurity: Strategic approaches to mitigate risk and safeguard data privacy. *World Journal of Advanced Research and Reviews*, 24(3), 310–327. <https://doi.org/10.30574/wjarr.2024.24.3.3695>
- [43] Nguyen, A., Ngo, H. N., Hong, Y., Dang, B., & Nguyen, B.-P. T. (2022). Ethical principles for artificial intelligence in education. *Education and Information Technologies*, 28(4), 4221–4241. <https://doi.org/10.1007/s10639-022-11316-w>
- [44] Overono, A. L., & Ditta, A. S. (2024). The Use of AI Disclosure Statements in Teaching: Developing Skills for Psychologists of the Future. *Teaching of Psychology*, 52(3), 273–278. <https://doi.org/10.1177/00986283241275664>
- [45] Perkins, M., Macvaugh, J., Roe, J., & Furze, L. (2024). The Artificial Intelligence Assessment Scale (AIAS): A Framework for Ethical Integration of Generative AI in Educational Assessment. *Journal of University Teaching and Learning Practice*, 21(06). <https://doi.org/10.53761/q3azde36>
- [46] Perkins, M., Roe, J., Vu, B. H., Postma, D., Hickerson, D., Mcgaughran, J., & Khuat, H. Q. (2024). Simple techniques to bypass GenAI text detectors: implications for inclusive education. *International Journal of Educational Technology in Higher Education*, 21(1). <https://doi.org/10.1186/s41239-024-00487-w>
- [47] Picht, P. G., & Thouvenin, F. (2023). AI and IP: Theory to Policy and Back Again – Policy and Research Recommendations at the Intersection of Artificial Intelligence and Intellectual Property. *IIC - International Review of Intellectual Property and Competition Law*, 54(6), 916–940. <https://doi.org/10.1007/s40319-023-01344-5>
- [48] Pitts, G., Rani, N., Mildort, W., & Cook, E.-M. (2025). Students' Reliance on AI in Higher Education: Identifying Contributing Factors. <https://doi.org/10.48550/arxiv.2506.13845>
- [49] Premkumar, P. P., Yatigammana, M. R. K. N., & Kannangara, S. (2024). Impact of Generative AI on Critical Thinking Skills in Undergraduates: A Systematic Review. *Journal of Desk Research Review and Analysis*, 2(1), 199–215. <https://doi.org/10.4038/jdr.v2i1.55>

- [50] Qaribilla, R., Mayawati, C. I., & Indrajaya, K. (2024). Digital Learning Inequality: The Role of Socioeconomic Status in Access to Online Education Resources. *International Journal of Social and Human*, 1(2), 51–58. <https://doi.org/10.59613/55gdm96>
- [51] Resnik, D. B., & Hosseini, M. (2025). Disclosing artificial intelligence use in scientific research and publication: When should disclosure be mandatory, optional, or unnecessary? *Accountability in Research*, ahead-of-print(ahead-of-print), 1–13. <https://doi.org/10.1080/08989621.2025.2481949>
- [52] Shahvaroughi Farahani, M., & Ghasemi, G. (2024). Artificial Intelligence and Inequality: Challenges and Opportunities. In *Qeios*. <https://doi.org/10.32388/7hwuz2>
- [53] Sharma, A. K., & Sharma, R. (2024). Generative Artificial Intelligence and Legal Frameworks: Identifying Challenges and Proposing Regulatory Reforms. *Kutafin Law Review*, 11(3), 415–451. <https://doi.org/10.17803/2713-0533.2024.3.29.415-451>
- [54] Sousa, A. E., & Cardoso, P. (2025). Use of Generative AI by Higher Education Students. *Electronics*, 14(7), 1258. <https://doi.org/10.3390/electronics14071258>
- [55] Spirgi, L. (2025). The Role of AI Disclosure in Academic Grading: Lecturer Perceptions, Challenges, and Implications (pp. 149–156). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-98465-5_19
- [56] Symeou, L., Louca, L., Kavadella, A., Mackay, J., Danidou, Y., & Raffay, V. (2025). Development of Evidence-Based Guidelines for the Integration of Generative AI in University Education Through a Multidisciplinary, Consensus-Based Approach. *European Journal of Dental Education*, 29(2), 285–303. <https://doi.org/10.1111/eje.13069>
- [57] Ugwu, N. F., Okorie, N. C., Adams, A. B., Igbinlade, A. S., Onayinka, T. S., Iroegbu, O., Opele, J. K., Ochiaka, R. E., Ojobola, F. B., Aigbona, P., Onyekwere, O. K., & Ezeani, U. D. (2024). Clarifying Ethical Dilemmas of Using Artificial Intelligence in Research Writing: A Rapid Review. *Higher Learning Research Communications*, 14(2). <https://doi.org/10.18870/hlrc.v14i2.1549>
- [58] Vieira, A., & Mesquita, A. (2025). Generative Artificial Intelligence in Higher Education: Challenges, Opportunities and Pedagogical Implications. *Journal of Technologies Information and Communication*, 5(1), 36578. <https://doi.org/10.55267/rtic/16675>
- [59] Wang, H., Dang, A., Wu, Z., & Mac, S. (2024). Generative AI in higher education: Seeing ChatGPT through universities' policies, resources, and guidelines. *Computers and Education: Artificial Intelligence*, 7, 100326. <https://doi.org/10.1016/j.caeai.2024.100326>
- [60] Wardat, Y. (2024). Exploring the Impact of ChatGPT on Scientific Research: Assessing Strengths, Weaknesses, Opportunities, and Threats. *Education as Change*, 28. <https://doi.org/10.25159/1947-9417/16006>
- [61] Weaver, K. (2024). The Artificial Intelligence Disclosure (AID) Framework: An Introduction. <https://doi.org/10.48550/arxiv.2408.01904>
- [62] Wen, H. (2024). Legal and Ethical Implications of AI-Generated Content in Intellectual Property Law. *Science of Law Journal*, 3(8). <https://doi.org/10.23977/law.2024.030802>
- [63] Xu, Z. (2025). Intellectual Property Rights Issues in AI-Generated Content. *Journal of Computer Technology and Electronic Research*, 2(7). <https://doi.org/10.70767/jcter.v2i7.741>
- [64] Yeter, I. H., Yang, W., & Sturgess, J. B. (2024). Global initiatives and challenges in integrating artificial intelligence literacy in elementary education: Mapping policies and empirical literature. *Future in Educational Research*, 2(4), 382–402. <https://doi.org/10.1002/fer3.59>
- [65] Yu, H., Hemphill, L., Li, L., Fan, L., Lee, S., & Ma, Z. (2023). ChatGPT in education: A discourse analysis of worries and concerns on social media. Cornell University. <https://doi.org/10.48550/arxiv.2305.02201>
- [66] Zhao, Y., Borelli, A., Martinez, F., Xue, H., & Weiss, G. M. (2024). Admissions in the age of AI: detecting AI-generated application materials in higher education. *Scientific Reports*, 14(1). <https://doi.org/10.1038/s41598-024-77847-z>