

# Beyond Plagiarism: Reconstructing Academic Integrity Framework under Conditions of Human-AI Co-Production

Agustinus Lanang Panji Cahyo <sup>1\*</sup>

<sup>1</sup> Driyarkara School of Philosophy, Jakarta, Indonesia

---

## Article Info

### Article history:

Received 26 February 2026

Revised 7 March 2026

Accepted 24 April 2026

---

### Keywords

Academic integrity

Generative artificial intelligence

Human-AI co-production

Higher education

Transparency




---

## Abstract

The rapid proliferation of generative artificial intelligence has fundamentally challenged traditional academic integrity frameworks, which have long centered on plagiarism as the primary form of misconduct. Unlike conventional copying, AI generated content is produced through probabilistic modeling and data synthesis, creating ambiguity in authorship, originality, and intellectual responsibility. This study examines the limitations of existing integrity frameworks in addressing the challenges of human and AI co production in higher education using a qualitative, literature based conceptual approach with thematic analysis of peer reviewed sources. The analysis identifies four key findings, namely the inadequacy of plagiarism centered frameworks, the rise of authorship ambiguity, the growing importance of transparency as an ethical principle, and the gap between institutional policies and actual academic practices. The findings indicate that ethical concerns are primarily linked to the absence of clear disclosure practices, resulting in forms of misconduct that are difficult to detect. In response, this study proposes a shift toward a transparency centered and collaboration aware framework, suggesting that educational institutions should integrate AI literacy into curricula, establish clear disclosure guidelines, and redesign assessment practices to emphasize critical thinking and process oriented evaluation

---

**Corresponding Author:** Cahyo, lanangpcahyo1@gmail.com

## 1. Introduction

The rapid proliferation of Artificial Intelligence (AI), particularly generative AI systems, has significantly transformed the landscape of higher education and academic knowledge production. In recent years, AI-powered tools have evolved from simple assistive technologies into sophisticated systems capable of generating complex, human-like academic texts. This transformation has fundamentally altered how students and researchers engage with writing, learning, and authorship. The scale of this shift is substantial, as a global survey conducted by the Digital Education Council found that 86 percent of students reported using AI tools in their studies, with more than half using them on at least a weekly basis (Digital Education Council, 2024). Such figures indicate that AI integration in academic settings is no longer a marginal or emerging phenomenon but a widespread and deeply embedded practice. Consequently, the traditional foundations of academic integrity, which have primarily been centered on the prevention of plagiarism, are increasingly strained by emerging forms of human and AI interaction and collaboration (Avello & Aranguren Zurita, 2025).

Historically, academic integrity has been framed through a relatively narrow lens, focusing on misconduct such as plagiarism, cheating, and unauthorized collaboration. Plagiarism, in particular, has long been treated as the central violation within academic institutions and is commonly defined as the act of presenting another individual's work as one's own without proper attribution. However, this conventional definition becomes increasingly problematic in the context of generative AI. Unlike traditional sources, AI-generated content is not directly copied from a single identifiable author but is instead produced through probabilistic modeling and large-scale data synthesis. As a result, questions of ownership, originality, and intellectual responsibility become increasingly ambiguous, thereby exposing the limitations of conventional plagiarism-based frameworks (Ebrahim et al., 2025).

Building on this shift, recent research indicates that the integration of AI into academic environments has grown at an unprecedented rate, reflecting both the opportunities and challenges posed by these

technologies. On one hand, AI has the potential to enhance learning, support writing processes, and improve accessibility in education. On the other hand, it introduces new ethical dilemmas, including the risk of over-reliance, diminished critical thinking, and the emergence of undetectable forms of academic misconduct. For instance, students may incorporate AI-generated outputs into their work without disclosure, raising concerns about transparency and fairness in assessment practices (Zhou & Wang, 2026). This concern is further reinforced by evidence suggesting that 58 percent of students have admitted to using AI tools to complete assignments dishonestly, underscoring the urgency of addressing these ethical concerns within institutional frameworks (Tan & Maravilla, 2024).

More fundamentally, the rise of human and AI co-production, where academic work is collaboratively produced by both human users and AI systems, has blurred the boundaries between assistance and authorship. In such contexts, it becomes increasingly difficult to determine the extent of human contribution in relation to machine-generated input. This growing ambiguity not only challenges institutional policies but also raises broader philosophical questions about the nature of knowledge creation and intellectual labor in the digital age. Consequently, existing integrity frameworks, which were originally designed for human-centered authorship, are increasingly inadequate in addressing these hybrid forms of production (Birks & Clare, 2023).

In response to these challenges, emerging scholarship suggests that maintaining academic integrity in the age of AI requires a shift away from purely surveillance-based approaches toward more adaptive and design-oriented strategies. Rather than focusing solely on detection and punishment, institutions are encouraged to develop policies that emphasize transparency, ethical AI usage, and assessment redesign. In this regard, integrity in AI-mediated environments is more effectively achieved through clear guidelines, disclosure practices, and pedagogical innovation, rather than reliance on detection technologies alone (Sangwa & Mutabazi, 2025).

Within this evolving context, the concept of academic integrity must be reconceptualized as a dynamic and adaptive framework rather than a static set of rules. It should account for the realities of human and AI collaboration while preserving core values such as honesty, accountability, and trust. This shift requires moving beyond plagiarism as the dominant paradigm toward a more comprehensive and context-sensitive understanding of ethical academic practice in technologically mediated environments.

Therefore, this study aims to critically examine the limitations of traditional academic integrity frameworks in addressing the challenges posed by human and AI co-production. Furthermore, it seeks to propose a reconstructed framework that integrates principles of transparency, responsibility, and ethical collaboration between humans and AI systems. By doing so, this research contributes to the ongoing discourse by offering a more adaptive and practically relevant approach to maintaining academic integrity in the evolving digital landscape.

## 2. Literature Review

The rapid advancement of Artificial Intelligence (AI), particularly generative AI, has significantly transformed the landscape of academic writing and integrity. In recent years, AI technologies have shifted from being supportive tools to becoming active contributors in knowledge production processes. (Avello & Aranguren Zurita, 2025) highlight that scholarly interest in AI and academic integrity has increased dramatically, reflecting the urgency of addressing the ethical and institutional implications of these technologies. However, while this growing body of literature demonstrates heightened academic attention, it also reveals a tendency to frame AI primarily as a disruptive force rather than as an integrated component of contemporary academic practice. This indicates that academic integrity can no longer be understood solely through traditional frameworks centered on plagiarism, but must instead be reconceptualized in response to evolving digital practices.

Traditional definitions of academic integrity have long emphasized plagiarism as the primary form of misconduct, defined as the act of presenting another's work as one's own. However, this definition becomes increasingly problematic in the context of AI-generated content. Unlike conventional plagiarism, AI does not replicate a single identifiable source but generates text through complex data synthesis and probabilistic modeling. The issue is therefore no longer limited to copying but extends to the lack of transparency in how AI contributes to the writing process. While this perspective successfully expands the scope of academic misconduct, it remains largely conceptual and does not fully address how such

transparency can be operationalized in real academic settings, highlighting a persistent gap between theoretical critique and practical implementation.

Expanding on this perspective, (Arar et al., 2025) argue that generative AI fundamentally alters the concept of authorship by introducing machine-generated contributions that are neither entirely human nor traditionally plagiarized. This creates a new category of knowledge production characterized by human and AI co-production, where the boundaries of intellectual ownership become increasingly blurred. However, although this argument effectively problematizes authorship, it does not provide clear criteria for distinguishing levels of human and AI contribution. As a result, the concept of authorship remains theoretically expanded but practically ambiguous, leaving educators and institutions without workable standards for evaluating the legitimacy of AI-assisted work.

The ethical implications of this shift are substantial. AI-assisted writing introduces challenges related to authenticity, accountability, and the integrity of intellectual labor. According to (Ilie, 2025), the integration of AI into academic practices raises critical concerns about whether students genuinely understand the content they submit or merely curate outputs generated by algorithms. While this concern is valid, it tends to assume passive student engagement and does not sufficiently consider the possibility of meaningful human and AI collaboration. This suggests that existing ethical discussions may be overly focused on risk, while underexploring the potential for responsible and productive AI integration. A more balanced ethical framework would need to account for both the dangers of over-reliance and the genuine intellectual value that thoughtful AI engagement can produce.

Empirical evidence further reveals a growing disconnect between institutional policies and actual student practices. Students frequently engage with AI tools in iterative and collaborative ways, often without explicit disclosure, suggesting that the primary issue is not the use of AI itself but the absence of clear guidelines and shared ethical standards (Sangwa & Mutabazi, 2025). Institutions, meanwhile, have been slow to develop policies that reflect this reality, resulting in a regulatory environment that is both inconsistent and insufficiently adaptive to the pace of technological change. This disconnect is not merely administrative but reflects a deeper normative uncertainty about what values academic integrity frameworks are meant to uphold in an era of human and AI collaboration.

Further reinforcing this point, (Zhou & Wang, 2026) highlight that AI-assisted practices can produce outputs that are difficult to detect using conventional plagiarism detection tools. This creates new forms of invisible misconduct, where violations are not easily identifiable through existing technological or institutional mechanisms. While this finding strongly supports the argument against detection-based approaches, it also raises an unresolved question regarding what alternative mechanisms can effectively replace them, a question that the existing literature has yet to answer with sufficient clarity or practical specificity.

Despite the growing body of literature, several critical gaps remain that this study seeks to address. The field currently lacks practical frameworks that can be directly implemented in academic institutions, as most existing work identifies the limitations of current systems without proposing concrete alternatives. Research on human and AI co-production also remains limited, with most studies continuing to frame AI either as a tool or as a threat rather than as a collaborative agent with distinct implications for authorship and accountability. Finally, the absence of sufficient empirical evidence on how students and researchers actually use AI in their writing processes constrains the development of evidence-based policies. In light of these gaps, there is a clear need for a reconstructed academic integrity framework that moves beyond traditional plagiarism-based approaches, recognizes the collaborative nature of human and AI knowledge production, and emphasizes transparency, accountability, and ethical responsibility as its foundational principles

### 3. Method

This study employs a qualitative research design with a conceptual and exploratory approach to investigate the transformation of academic integrity frameworks under conditions of human and AI co-production. Rather than conducting a purely systematic review, this study adopts a conceptual synthesis approach aimed at developing a theoretical framework grounded in existing literature. This approach is particularly suitable for examining complex ethical and conceptual issues that cannot be fully captured through quantitative methods, as it prioritizes depth of interpretation over breadth of generalization. (Birks & Clare, 2023) emphasize that AI-assisted academic practices introduce nuanced forms of misconduct that

require in-depth qualitative exploration, further supporting the appropriateness of this methodological orientation.

The research is structured as a conceptual and literature-informed study that aims to critically evaluate existing academic integrity frameworks and develop a reconstructed model in response to the challenges posed by human and AI co-production. As noted by O'Dea & O'Dea (2023), the rapid development of AI in higher education requires conceptual approaches to understand its broader implications before empirical generalizations can be established. Accordingly, this study focuses on synthesizing existing knowledge to identify conceptual gaps, interrogate theoretical assumptions, and propose adaptive solutions that are responsive to the current technological landscape.

The study draws on a range of peer-reviewed and open-access academic publications relevant to artificial intelligence, academic integrity, and human and AI collaboration. Sources were selected from Google Scholar and DOAJ with priority given to publications from 2023 onwards to reflect the most current developments in the field. Institutional policies and ethical guidelines were also examined to provide practical perspectives on how academic integrity is currently addressed in AI-mediated contexts. This combination of theoretical and institutional sources allows the study to bridge conceptual insights with real-world practices, strengthening the practical applicability of the proposed framework.

The analysis is conducted using a thematic synthesis approach to identify recurring patterns and key concepts across the selected literature. This process involves several interconnected stages, including data familiarization, open coding, theme development, and interpretive synthesis. Key concepts such as authorship ambiguity, AI misuse, and transparency are identified and grouped into broader thematic categories, which subsequently form the basis for understanding the limitations of existing frameworks and informing the development of a new conceptual model. AI-assisted practices often produce forms of misconduct that are difficult to detect using conventional mechanisms, reinforcing the need for deeper conceptual analysis rather than surface-level policy responses (Ouyang et al., 2022; Sangwa & Mutabazi, 2025).

To ensure the credibility and trustworthiness of the study, several strategies are employed throughout the research process. The use of diverse and credible academic sources enhances the robustness of the analysis, while the integration of multiple theoretical perspectives supports conceptual depth and reduces the risk of interpretive bias. The study also applies a form of triangulation by comparing findings across different types of sources, including peer-reviewed academic literature and institutional policy documents. Combining multiple perspectives strengthens the trustworthiness of qualitative research, particularly in interdisciplinary fields involving rapidly evolving technologies such as AI. This multi-source approach ensures that the conclusions drawn are not overly dependent on any single body of literature, thereby enhancing the overall validity of the proposed framework.

#### **4. Results**

The analysis of the selected literature reveals several key findings that illustrate the transformation of academic integrity frameworks in the context of human and AI co-production. Through thematic synthesis, major patterns are identified, namely the inadequacy of plagiarism-based frameworks, the rise of authorship ambiguity, the centrality of transparency, and the gap between institutional policies and academic practices. In addition, the analysis highlights the emergence of invisible misconduct and the limited availability of empirical evidence, both of which further complicate the development of effective academic integrity models.

The first and most foundational finding concerns the inadequacy of traditional plagiarism-based frameworks in addressing contemporary academic practices. AI-generated content does not align with conventional definitions of plagiarism, as it is produced through probabilistic systems rather than direct copying. As highlighted by Ebrahim et al., (2025), this shift disrupts the binary distinction between original and non-original work, thereby exposing the limitations of existing integrity models. This limitation becomes more apparent in academic contexts where AI-generated content is integrated into writing processes in iterative and collaborative ways. Rather than copying existing sources, users often engage in co-creation with AI systems, making it difficult to categorize such practices within existing definitions of misconduct. Plagiarism-based frameworks therefore fail to capture the complexity of AI-mediated academic production, rendering them insufficient as the primary basis for institutional integrity policies.

Closely related to this is the second finding, which concerns the increasing ambiguity of authorship in AI-mediated academic production. In human and AI co-production contexts, it becomes difficult to determine the extent of human contribution relative to machine-generated input, challenging conventional notions of intellectual ownership and complicating attribution practices within academic writing. The absence of clear guidelines regarding authorship attribution in AI-assisted work intensifies this issue considerably. Without standardized criteria, academic institutions struggle to define what constitutes legitimate authorship, leading to inconsistencies in evaluation and ethical judgment. This ambiguity highlights the need for more nuanced frameworks that can accommodate hybrid forms of knowledge production while maintaining meaningful standards of accountability.

The third finding concerns the emergence of transparency as a central principle in redefining academic integrity. The literature suggests that ethical concerns are less related to the use of AI itself and more closely associated with how such use is disclosed. Sangwa and Mutabazi (2025) emphasize that the absence of disclosure practices can lead to misrepresentation of authorship and undermine academic trust. In this context, transparency functions as a mechanism for maintaining accountability in AI-mediated academic work, as clear disclosure of AI involvement allows educators and institutions to better assess the extent of human contribution and support fair evaluation practices. This finding indicates that transparency-based approaches may offer a more adaptive and practically viable alternative to traditional detection-based models.

The fourth finding identifies a significant gap between institutional policies and actual academic practices. While many institutions continue to rely on traditional plagiarism detection systems, students are already integrating AI tools into their academic workflows in complex and often undocumented ways. Avello and Aranguren Zurita (2025) note that institutional responses tend to be reactive rather than adaptive, limiting their effectiveness in addressing emerging challenges. This misalignment creates a situation in which existing policies fail to reflect the realities of academic behavior, increasing the likelihood of unintentional misconduct among students who operate within unclear or outdated guidelines. The gap underscores the urgent need for policy frameworks that are responsive to technological developments and genuinely aligned with actual user practices.

Tabel 1. Comparison of Traditional and Proposed Academic Integrity Frameworks

<b>Dimension</b>	<b>Traditional Framework</b>	<b>Proposed Framework</b>
Core concept	Plagiarism detection	Ethical transparency
View of AI	Threat/violation tool	Collaborative agent
Authorship model	Individual ownership	Contribution-based
Misconduct type	Visible copying	Invisible misconduct
Institutional response	Reactive, Punitive	Adaptive, educative
Evaluation focus	Textual similarity	Process and intent
Key mechanism	Detection technology	Disclosure and reflection

Beyond these four primary findings, the analysis also surfaces additional concerns that further complicate the integrity landscape. The first is the emergence of invisible misconduct, whereby AI-assisted writing produces outputs that are difficult to detect using conventional tools. Chaka (2023), Ibrahim (2026), and Tsigaris & Teixeira da Silva (2026) demonstrate that AI-generated or AI-assisted content often bypasses detection systems entirely, representing a shift from visible to less detectable forms of academic misconduct and rendering surveillance-based enforcement increasingly unreliable. The second is the persistent lack of empirical evidence on actual AI usage in academic contexts. This conceptual gap constrains the development of practical and implementable frameworks, since without empirical insights into how students and researchers interact with AI tools, it becomes difficult to design policies that are both effective and contextually relevant (Derin et al., 2025).

Taken together, these findings demonstrate a fundamental shift from a plagiarism-centered model toward a more complex and adaptive understanding of academic integrity that must account for the realities of human and AI collaboration. The identified themes collectively provide a coherent foundation for the development of a reconstructed academic integrity framework that emphasizes transparency, accountability, and ethical responsibility as its guiding principles in AI-mediated academic environments (Gray et al., 2025; Nguyen & Goto, 2024).

## 5. Discussion

The findings of this study point to a structural transformation in academic integrity discourse, one that demands more than incremental policy adjustment. While the inadequacy of plagiarism-based frameworks is now widely acknowledged in the literature, the field risks stalling at diagnosis without arriving at a credible prescription. This discussion critically interrogates the implications of the findings, challenges several assumptions embedded in current scholarship, and proposes directions that move beyond merely identifying the problem.

The declining centrality of plagiarism as the primary marker of academic misconduct is, as Ebrahim et al. (2025) suggest, a logical consequence of how AI-generated content is produced. Unlike conventional plagiarism, which involves the appropriation of an identifiable source, AI output emerges from probabilistic synthesis across vast datasets, a process that disrupts the binary logic of "original" versus "copied." This shift in understanding is necessary, but it should not be treated as sufficient. What the literature has been slower to address is the risk of overcorrection. In moving away from plagiarism-centered thinking, some scholars implicitly treat transparency as a straightforward replacement, as though disclosure alone resolves the ethical complexity of AI-assisted work. This assumption deserves scrutiny. Transparency is a procedural mechanism, not an ethical guarantee. A student may fully disclose the use of AI and still produce work that represents minimal intellectual engagement. Conversely, a student who uses AI iteratively as a thinking tool may contribute deeply to the intellectual process while struggling to articulate the nature of that contribution in any disclosure statement. Transparency matters, but it cannot bear the full moral weight currently assigned to it.

The reconceptualization of authorship in human-AI co-production contexts, moving from an individual ownership-based model toward a contribution-based framework, represents a genuine theoretical advance. Arar et al. (2025), as synthesized in this study, effectively destabilize the assumption that authorship is a singular and clearly bounded act. Yet this theoretical move arrives without the criteria needed to operationalize it. The fundamental question of how much human contribution is sufficient to constitute legitimate authorship in an AI-assisted context remains unanswered. This is not merely an administrative inconvenience; it is a foundational gap that undermines any proposed framework. Without workable criteria, the concept of contribution-based authorship risks becoming a flexible standard that can be invoked selectively, privileging students who are more skilled at articulating their process over those who actually engaged more deeply. This points to a tension between equity and accountability that the existing literature has yet to adequately address. Tan and Maravilla (2024) further complicate this picture by noting that for AI use in education to be ethically responsible, it must adhere to principles of fairness and accountability, suggesting that any authorship framework must be evaluated not only for its conceptual coherence but also for its equitable application across diverse student populations.

Zhou and Wang (2026) demonstrate that AI-assisted writing increasingly bypasses conventional detection systems, producing what this study terms invisible misconduct. This finding is empirically significant, but its implications are often drawn too quickly. The standard conclusion that detection-based approaches are insufficient is correct, yet the proposed alternative of transparency and ethical awareness may not be as robust as it appears. The assumption underlying transparency-based models is that students and institutions share a common ethical vocabulary for what constitutes acceptable AI use. Evidence suggests otherwise. As Avello and Aranguren Zurita (2025) observe, institutional responses have been reactive and inconsistent, and students frequently operate under ambiguous or contradictory guidelines. In the absence of shared norms, calls for transparency place the moral burden on individuals navigating a space their institutions have not yet defined. This is not merely a policy lag; it is a structural condition that transparency alone cannot resolve. Effective reform requires not only disclosure requirements but the prior construction of shared normative standards, a task that is considerably more complex and contested.

The identified gap between institutional policies and actual academic practices is often framed as a failure of institutional responsiveness. While this characterization is partially valid, it risks misdiagnosing the underlying problem. The gap is not simply a matter of policies failing to keep pace with technological change; it reflects a deeper uncertainty about what values academic integrity frameworks are actually meant to protect. If the core purpose of academic integrity is to ensure that assessment accurately reflects individual learning and capability, then human-AI co-production poses a fundamental challenge regardless of how transparent the process is. If, on the other hand, the purpose is to foster honest engagement with knowledge production however mediated, then the ethical stakes shift considerably. Current debates

frequently conflate these two purposes, producing policy frameworks that satisfy neither. A reconstructed academic integrity model must therefore begin by resolving this normative ambiguity, not by assuming it away. As Tan & Maravilla, (2024) argue, the field currently lacks a coherent theoretical foundation capable of guiding institutional responses, with most existing work remaining descriptive rather than prescriptive in orientation.

The findings collectively support the case for a transparency-centered and collaboration-aware framework. However, the conditions under which such a framework can function effectively must be stated clearly. First, it requires the development of discipline-specific norms for AI disclosure, since the ethical implications of AI assistance differ substantially across fields such as law, medicine, creative writing, and engineering. Second, it requires assessment redesign that shifts evaluation toward process documentation, iterative drafts, and oral examination, not simply the addition of disclosure statements to existing submission formats. In this regard, Khlaif et al., (2025) demonstrate that redesigning assessments to emphasize critical thinking and process-oriented evaluation represents one of the most effective strategies for maintaining academic integrity in AI-mediated environments, as such designs inherently reduce the value of AI-generated outputs by foregrounding authentic intellectual labor. Third, and most critically, it requires institutional honesty about the limits of any integrity framework. No model, whether traditional or reconstructed, can eliminate the possibility of undetectable misconduct. What a well-designed framework can do is create conditions in which honest engagement is structurally supported and ethically incentivized, rather than merely formally required. The shift away from plagiarism as the organizing metaphor of academic integrity is overdue, but moving beyond plagiarism is only the first step. The harder task, one that the existing literature has not yet fully undertaken, is constructing the normative and institutional architecture that makes a new framework not only theoretically coherent but practically livable for the students and educators it is meant to serve.

Building on these preconditions, this study proposes a reconstructed academic integrity framework organized around three core principles. The first is ethical transparency, which moves beyond simple disclosure requirements toward a culture of reflective documentation, in which students are expected to articulate not only whether AI was used, but how it shaped their thinking and what intellectual contributions they made independently. The second is collaborative accountability, which recognizes human-AI co-production as a legitimate mode of knowledge work while establishing that the human author retains full responsibility for the reasoning, judgment, and conclusions expressed in the final work. The third is contextual adaptability, which acknowledges that no single standard can govern AI use across all disciplines and assessment types, and that institutions must therefore develop field-specific guidelines that reflect the distinct epistemic values of each domain. Together, these principles constitute a framework that is neither permissive nor prohibitive, but rather oriented toward fostering genuine intellectual engagement in an environment where AI is an increasingly unavoidable presence. Rather than positioning integrity as a mechanism of control, this framework reconceives it as a shared commitment among students, educators, and institutions to preserve the conditions under which meaningful learning and honest knowledge production remain possible.

In response to these challenges, this study proposes a shift from a plagiarism-centered paradigm toward a transparency-centered and collaboration-aware framework. By emphasizing ethical disclosure, responsible AI use, and contextual evaluation, this approach offers a more realistic and sustainable model for maintaining academic integrity in AI-mediated environments. Critically, this shift is not merely conceptual but carries direct implications for how institutions design assessments, train educators, and communicate expectations to students.

From a practical perspective, academic institutions are encouraged to move beyond reliance on detection technologies and adopt more comprehensive strategies, including the integration of AI literacy, the establishment of clear disclosure policies, and the redesign of assessment practices. Such measures are essential to ensure that integrity frameworks remain relevant and effective in the face of rapid technological change. Ultimately, the evolution of artificial intelligence necessitates a fundamental rethinking of academic integrity. Moving beyond plagiarism is not merely an option, but a necessity to ensure that academic systems remain credible, ethical, and adaptable in the era of human and AI collaboration.

Despite its contributions, this study is subject to several important limitations that should be acknowledged. First, the conceptual and literature-based nature of this research means that the proposed

framework has not been empirically tested in real academic settings. The findings are therefore theoretical in orientation and cannot be generalized to specific institutional contexts without further validation. Second, the study relies on a selected body of peer-reviewed literature rather than a systematic review with explicit inclusion and exclusion criteria, which introduces the possibility of selection bias in the sources consulted. Third, the majority of literature drawn upon reflects perspectives from higher education contexts in the Global North, which may limit the applicability of the proposed framework to educational systems with different cultural, institutional, and technological conditions.

Future research should address these limitations by pursuing empirical investigations into how students and educators actually engage with AI tools across different academic disciplines and institutional settings. Longitudinal studies tracking changes in AI use practices over time would be particularly valuable in capturing the dynamic nature of this phenomenon. Additionally, interdisciplinary approaches that integrate perspectives from education, ethics, philosophy, and technology studies are needed to develop more comprehensive and nuanced integrity frameworks. Comparative and cross-cultural research would also contribute meaningfully to this field, as the ethical norms surrounding AI use in academic contexts are likely to vary significantly across different national and institutional environments. Finally, future work should explore the practical implementation of transparency-based frameworks, examining how disclosure policies and assessment redesign function in real educational contexts and what barriers exist to their adoption. Such research would provide the empirical grounding necessary to transform the conceptual model proposed in this study into actionable institutional policy. This reinforces the urgency of developing academic integrity frameworks that are not only theoretically robust but also practically sustainable in rapidly evolving educational environments.

## 6. Conclusion

This study demonstrates that the integration of artificial intelligence into academic practices has fundamentally reshaped the foundations of academic integrity. Traditional frameworks centered on plagiarism detection are no longer sufficient to address the complexities introduced by human and AI co-production, where boundaries of authorship, originality, and intellectual responsibility are increasingly blurred. The core assumption that misconduct can be identified through textual similarity has been rendered inadequate by the probabilistic and generative nature of AI systems, which produce content that is neither copied nor wholly original in any conventional sense. The findings highlight that the central challenge lies not in the use of AI itself, but in the absence of transparency and clear ethical guidelines governing its use. The emergence of authorship ambiguity and invisible misconduct underscores the limitations of detection-based approaches and supports the need for more adaptive and principle-driven models. Without a shared normative foundation, students and institutions are left navigating an ethically ambiguous landscape in which the rules of engagement remain unclear and inconsistently enforced.

## References

- Arar, K. H., Özen, H., Polat, G., & Turan, S. (2025). Artificial intelligence, generative artificial intelligence and research integrity: A hybrid systemic review. *Smart Learning Environments*, 12(1), 44. <https://doi.org/10.1186/s40561-025-00403-3>
- Avello, D., & Aranguren Zurita, S. (2025). Exploring the nexus of academic integrity and artificial intelligence in higher education: A bibliometric analysis. *International Journal for Educational Integrity*, 21(1), 24. <https://doi.org/10.1007/s40979-025-00199-2>
- Birks, D., & Clare, J. (2023). Linking artificial intelligence facilitated academic misconduct to existing prevention frameworks. *International Journal for Educational Integrity*, 19(1), 20. <https://doi.org/10.1007/s40979-023-00142-3>

- Chaka, C. (2023). Detecting ai content in responses generated by ChatGPT, Youchat, and Chatsonic: The case of five ai content detection tools. *Journal of Applied Learning & Teaching*, 6(2), 94–104. <https://doi.org/10.3316/informit.T2025102600002400418942396>
- Derin, T., Hartati, P. T., Hamuddin, B., & Abbas, M. F. F. (2025). "Do you even know what AI is?": Assessing Students' AI Awareness to Realize a Critical Digital Pedagogy. *Jurnal Pendidikan Tambusai*, 9(3), 41763–41770. <https://doi.org/10.31004/jptam.v9i3.35682>
- Digital Education Council. (2024). *Global AI Student Survey 2024*. <https://www.digitaleducationcouncil.com/post/what-students-want-key-results-from-dec-global-ai-student-survey-2024>
- Ebrahim, O. O., Mhouti, A. E., & Allaoui, M. (2025). Academic Integrity in the Era of Artificial Intelligence: A Systematic Review. In M. Khaldi, M. Kaddouri, & M. Erradi (Eds.), *Proceedings of the E-Learning and Smart Engineering Systems (ELSEES 2024)* (Vol. 36, pp. 99–113). Atlantis Press SARL. [https://doi.org/10.2991/978-2-38476-408-2\\_9](https://doi.org/10.2991/978-2-38476-408-2_9)
- Gray, S. L., Edsall, D., & Parapadakis, D. (2025). AI-Based Digital Cheating At University, and the Case for New Ethical Pedagogies. *Journal of Academic Ethics*, 23(4), 2069–2086. <https://doi.org/10.1007/s10805-025-09642-y>
- Ibrahim, K. (2026). Using AI to control AI-assisted plagiarism in L2 writing: A critical review of the literature. *Language Testing in Asia*, 16(1), 21. <https://doi.org/10.1186/s40468-026-00433-9>
- Ilie, O.-A. (2025). The Ethics of AI-Assisted Academic Writing. Authenticity Criteria in the Evaluation of Students' Assignments. *International Conference KNOWLEDGE-BASED ORGANIZATION*, 31(2), 155–158. <https://doi.org/10.2478/kbo-2025-0063>
- Khlaif, Z. N., Alkouk, W. A., Salama, N., & Abu Eideh, B. (2025). Redesigning Assessments for AI-Enhanced Learning: A Framework for Educators in the Generative AI Era. *Education Sciences*, 15(2), 174. <https://doi.org/10.3390/educsci15020174>
- Nguyen, H. M., & Goto, D. (2024). Unmasking academic cheating behavior in the artificial intelligence era: Evidence from Vietnamese undergraduates. *Education and Information Technologies*, 29(12), 15999–16025. <https://doi.org/10.1007/s10639-024-12495-4>

- O'Dea, X., & O, M. (2023). Is artificial intelligence really the next big thing in learning and teaching in higher education?: A conceptual paper. *Journal of University Teaching and Learning Practice*, 20(5), 1–17. <https://doi.org/10.3316/informit.T2024112700021091343163252>
- Ouyang, F., Zheng, L., & Jiao, P. (2022). Artificial intelligence in online higher education: A systematic review of empirical research from 2011 to 2020. *Education and Information Technologies*, 27(6), 7893–7925. <https://doi.org/10.1007/s10639-022-10925-9>
- Sangwa, S., & Mutabazi, P. (2025). *Generative AI and Academic Integrity in Online Distance Learning: The AI-Aware Assessment Policy Index for the Global South*. Social Sciences. <https://doi.org/10.20944/preprints202509.2460.v1>
- Tan, M. J. T., & Maravilla, N. M. A. T. (2024). Shaping integrity: Why generative artificial intelligence does not have to undermine education. *Frontiers in Artificial Intelligence*, 7, 1471224. <https://doi.org/10.3389/frai.2024.1471224>
- Tsigaris, P., & Teixeira da Silva, J. A. (2026). AI detecting AI in academic writing: Why most AI detector findings are false. *Next Research*, 7, 101396. <https://doi.org/10.1016/j.nexres.2026.101396>
- Zhou, X., & Wang, X. (2026). Upholding academic integrity: An exploratory study of AI-assisted detection of unauthorised machine translation use in student translations. *Humanities and Social Sciences Communications*, 13(1), 331. <https://doi.org/10.1057/s41599-026-06827-7>