

Generative Artificial Intelligence and Academic Integrity in Higher Education: A Critical Policy Review

Muhammad Ikhlas Rosele¹Muhammad Saidal Arifin Harahap²

¹ Ketua Jabatan Fiqh & Usul, Akademi Pengajian Islam (APIUM), Universiti Malaya, Kuala Lumpur

²PTKU MUI Sumatera Utara

Email: ikhlas@um.edu.my¹ arifinsaid41h@gmail.com²

Article History	Abstract
Received : 08-13-2026 Revised : 09-18-2026 Published : 10-01-2026	The public availability of generative artificial intelligence tools since early 2023 has disrupted the assumptions on which academic integrity regimes in higher education were built, since those regimes were designed to detect the appropriation of existing text rather than the generation of novel text on demand. This article aims to review critically how the scholarly and policy response has developed, to evaluate the three dominant institutional strategies of detection, prohibition, and assessment redesign, and to assess the adequacy of the Indonesian regulatory framework for governing academic conduct under these conditions. The study is a critical, non-systematic review combining thematic synthesis of peer-reviewed literature published between 2023 and 2025 with document analysis of Indonesian regulations on plagiarism and higher education quality assurance; it does not report original screening statistics and makes no claim to exhaustive coverage. The review finds that the discourse has shifted from containment toward pedagogical redesign, that detection-based strategies rest on an evidential basis too weak to sustain disciplinary decisions, that blanket prohibition produces unenforceable rules and inequitable outcomes, and that assessment redesign is the only strategy whose effectiveness does not depend on distinguishing human from machine text. It also finds that the Indonesian framework, constructed around plagiarism understood as the appropriation of another person's work, does not straightforwardly reach undisclosed machine generation. Institutions are advised to regulate disclosure rather than use, to design assessment for process evidence, and to treat artificial intelligence literacy as a graduate competence.
Keywords: Academic integrity; Assessment design; Generative artificial intelligence; Higher education policy; Research ethics.	

I. INTRODUCTION

Academic integrity regimes in higher education were constructed around a specific assumption about how misconduct occurs. Their central prohibition, plagiarism, presupposes an antecedent text belonging to an identifiable author, and their central detection technology, textual similarity matching, presupposes that appropriated language will resemble a source held in a searchable corpus.¹ Indonesian regulation reflects this architecture precisely: the ministerial regulation on the prevention of plagiarism defines the offence as obtaining credit for a scholarly work by quoting part or all of another party's work without adequate attribution, and the Higher Education Law attaches degree revocation to proven plagiarism.²

¹Tracey Bretag, ed., *Handbook of Academic Integrity* (Singapore: Springer, 2016), 1–14.

²Regulation of the Minister of National Education Number 17 of 2010 concerning the Prevention and Handling of Plagiarism in Higher Education, Article 1(1); Law Number 12 of 2012 concerning Higher Education, Article 70.

Generative artificial intelligence violates that assumption. Text generated on demand has no antecedent author and no corpus counterpart, so the conduct at issue is not appropriation but undisclosed delegation of authorship. Since early 2023, when such tools became widely available, higher education institutions have responded with an unusual combination of speed and improvisation, issuing prohibitions, procuring detection services, and revising assessment policies, often within a single academic year.³ The scholarly response has been correspondingly rapid and, in its early phase, heavily oriented toward containment.⁴

Three years of accumulated literature now permit a more considered assessment. Systematic reviews have mapped the landscape of generative artificial intelligence in higher education and documented both the pedagogical opportunities claimed for it and the integrity risks attributed to it.⁵ Empirical work has tested the central premise of detection-based strategies by examining whether experienced educators can identify machine-generated student writing, with results that are not encouraging for that premise.⁶ Reviews of attitudes and utilisation behaviours have shown that student adoption has substantially outpaced institutional policy.⁷ Commentary has moved from questions of prohibition toward questions of assessment design and artificial intelligence literacy.⁸

What remains underdeveloped is a critical evaluation of the strategies themselves against the regulatory conditions of a particular national system. Most policy discussion is either generic or anchored in jurisdictions whose integrity regimes differ materially from Indonesia's. The research gap this article addresses is therefore twofold: the absence of a strategy-by-strategy evaluation grounded in the evidential requirements of disciplinary decision-making, and the absence of an assessment of whether the Indonesian framework, built around plagiarism as appropriation, reaches the conduct now at issue.

The objectives are to review critically how the scholarly and policy response has developed since 2023, to evaluate the strategies of detection, prohibition, and assessment redesign, and to assess the adequacy of the Indonesian regulatory framework. Three research questions follow: how has the discourse on generative artificial intelligence and academic integrity developed; what are the evidential and practical limits of each dominant institutional strategy; and does the Indonesian regulatory framework provide an adequate basis for governing undisclosed machine generation?

³Jürgen Rudolph, Samson Tan, and Shannon Tan, "ChatGPT: Bullshit Spewer or the End of Traditional Assessments in Higher Education?," *Journal of Applied Learning and Teaching* 6, no. 1 (2023): 342–363.

⁴Debby R. E. Cotton, Peter A. Cotton, and J. Reuben Shipway, "Chatting and Cheating: Ensuring Academic Integrity in the Era of ChatGPT," *Innovations in Education and Teaching International* 61, no. 2 (2024): 228–240.

⁵Ahmed Tlili et al., "Mapping the Landscape of Generative AI in Education: A Systematic Review," *Computers and Education: Artificial Intelligence* 5 (2024): 100167.

⁶Johanna Fleckenstein et al., "Do Teachers Spot AI? Evaluating the Detectability of AI-Generated Texts among Student Essays," *Computers and Education: Artificial Intelligence* 6 (2024): 100209.

⁷Fan Wu, Yang Dang, and Manli Li, "A Systematic Review of Responses, Attitudes, and Utilization Behaviors on Generative AI for Teaching and Learning in Higher Education," *Behavioral Sciences* 15, no. 4 (2025): 467, <https://doi.org/10.3390/bs15040467>.

⁸UNESCO, *Guidance for Generative AI in Education and Research* (Paris: UNESCO, 2023), 8–20.

II. RESEARCH METHODS

This study is a critical, interpretive review rather than a systematic review. The distinction is stated explicitly because the two designs make different claims: a systematic review asserts exhaustive and reproducible coverage of a defined corpus according to a pre-registered protocol, whereas a critical review selects purposively in order to evaluate arguments and identify conceptual weaknesses.⁹ This article makes the latter claim only. It reports no screening statistics and does not assert that its coverage is exhaustive; readers seeking comprehensive mapping are directed to the systematic reviews cited below, on whose coverage this article partly relies.

Sources were selected purposively in three categories. The first comprises peer-reviewed reviews and empirical studies published between 2023 and 2025 addressing generative artificial intelligence in higher education, academic integrity, or assessment. The second comprises the established literature on academic integrity and assessment security that predates the current disruption, included because the evaluation of new strategies requires criteria drawn from settled scholarship rather than from the disrupted present. The third comprises Indonesian regulatory instruments governing plagiarism, higher education quality assurance, and research ethics.

Analysis proceeded by thematic synthesis. Sources were read and coded for claims about the nature of the integrity problem, for claims about the efficacy of particular institutional responses, and for the evidential basis offered for those claims. Codes were grouped into the three strategies that structure the discussion, and each strategy was then evaluated against three criteria: evidential sufficiency, meaning whether the strategy can generate evidence capable of supporting a disciplinary decision; enforceability, meaning whether compliance can be monitored without disproportionate intrusion; and equity, meaning whether the strategy imposes differential burdens on identifiable groups of students. The Indonesian regulatory analysis then applied ordinary techniques of legal interpretation to determine the reach of existing definitions. The principal limitation is that a rapidly moving field will have generated relevant work after the period reviewed.

III. RESULTS AND DISCUSSION

A. From Containment to Redesign: The Trajectory of the Discourse

The first finding concerns the development of the discourse. Publications from the first phase, concentrated in 2023, are predominantly oriented toward containment: they discuss detection, prohibition, and the mitigation of misconduct, and they frame the technology principally as a threat to assessment validity.¹⁰ By 2024 and 2025 the centre of gravity has shifted toward pedagogical redesign, the integration of artificial intelligence literacy into curricula, and the governance of legitimate use.¹¹ This trajectory is not merely a change of fashion; it reflects the accumulation of evidence that containment

⁹Maria J. Grant and Andrew Booth, "A Typology of Reviews: An Analysis of 14 Review Types and Associated Methodologies," *Health Information and Libraries Journal* 26, no. 2 (2009): 91–108.

¹⁰Rudolph, Tan, and Tan, "ChatGPT," 345–352; Cotton, Cotton, and Shipway, "Chatting and Cheating," 230–235.

¹¹Tlili et al., "Mapping the Landscape of Generative AI in Education"; Wu, Dang, and Li, "A Systematic Review of Responses, Attitudes, and Utilization Behaviors."

strategies do not work, and it follows a pattern familiar from earlier technological disruptions of assessment.

A second feature of the trajectory deserves comment. Much of the early literature treated the problem as one of student honesty, and proposed interventions accordingly directed at student conduct. The later literature increasingly locates the problem in assessment design, on the reasoning that an assessment which can be completed satisfactorily by a machine was already measuring something other than the competence it purported to measure.¹² This reframing has an uncomfortable implication that institutions have been slow to accept: the disruption has exposed a pre-existing weakness rather than created a new one.

B. The Three Strategies Evaluated

The second finding concerns the comparative merits of the dominant strategies, evaluated against evidential sufficiency, enforceability, and equity.

Detection. Strategies premised on identifying machine-generated text fail the evidential sufficiency criterion. Empirical testing indicates that educators, whether novice or experienced, are unreliable at distinguishing machine-generated from student-written text.¹³ Automated detection services report probabilistic outputs whose error characteristics are not independently validated for the populations to which they are applied, and false positives fall unevenly, with particular concern reported for writing by students using a second language. A disciplinary decision requires evidence capable of supporting a finding to the applicable standard of proof; a probability score of unknown calibration does not meet that requirement, and institutions that treat it as though it does expose themselves to justified challenge. Detection may retain a legitimate diagnostic role in prompting conversation, but it cannot bear evidential weight.

Prohibition. Blanket prohibition fails principally on enforceability and equity. It cannot be monitored without the detection capability whose limits have just been described, so the prohibition's operative effect is to sanction those who are careless rather than those who are dishonest, which inverts the culpability the rule is meant to track. It also generates a secondary harm: where legitimate uses, such as language support or idea generation, are prohibited alongside illegitimate ones, students who would have disclosed use are given an incentive to conceal it, and the institution loses the information on which any constructive response depends. The comparative literature on integrity has long established that rules perceived as arbitrary reduce rather than increase compliance.¹⁴

Assessment redesign. Redesign is the only strategy whose effectiveness does not depend on distinguishing human from machine production, because it changes what is assessed rather than attempting to police how it was produced. Its instruments are established independently of the current disruption: assessment of process as well as

¹²Phillip Dawson, *Defending Assessment Security in a Digital World: Preventing E-Cheating and Supporting Academic Integrity in Higher Education* (London: Routledge, 2021), 3–18.

¹³Fleckenstein et al., “Do Teachers Spot AI?,” 3–8.

¹⁴Donald L. McCabe, Kenneth D. Butterfield, and Linda K. Treviño, *Cheating in College: Why Students Do It and What Educators Can Do about It* (Baltimore: Johns Hopkins University Press, 2012), 60–75.

product, requirement of evidence of intermediate reasoning, oral defence or interactive components, and tasks anchored in specific contexts that the assessor can verify.¹⁵ The literature on assessment security identifies this as the structurally sound response, on the reasoning that security should be designed into the assessment rather than added as surveillance afterwards.¹⁶ Its cost is real, since redesign is labour-intensive and oral components are resource-intensive at scale, and any honest evaluation must acknowledge that this cost is the principal obstacle to adoption.

Table 1. Comparative Evaluation of Institutional Strategies

Strategy	Evidential sufficiency	Enforceability	Equity
Detection	Low: probabilistic, uncalibrated	Low	Low: uneven false positives
Blanket prohibition	Dependent on detection	Low	Low: penalises disclosure
Disclosure requirement	Moderate: shifts burden to declaration	Moderate	High
Assessment redesign	High: assesses observable process	High	High

Source: Thematic synthesis of reviewed literature, 2026.

Table 1 supports a clear ordering. The two strategies that dominated the early institutional response are the two that perform worst on every criterion, while the strategy that performs best is the most demanding to implement. This is a recognisable pattern in regulatory design, in which the response that is cheapest to announce is inversely related to the response that works.

C. The Indonesian Regulatory Framework

The third finding concerns Indonesian regulation. The ministerial regulation on plagiarism defines the offence in terms of obtaining credit for a scholarly work by using part or all of another party's work or ideas without adequate and sufficient attribution.¹⁷ Interpreted according to ordinary principles, the definition presupposes a work belonging to another party. Text generated by a model does not straightforwardly answer to that description: it is not the work of another party in the sense the regulation contemplates, and the conduct at issue is the concealment of how the work was produced rather than the appropriation of someone else's expression.

Two consequences follow. First, an institution that disciplines a student for undisclosed machine generation under the plagiarism provision is applying a norm by analogy, which is defensible in internal academic governance but vulnerable where the sanction is severe, and degree revocation for proven plagiarism is provided by statute.¹⁸ Second, the gap is best closed at the level at which the conduct actually occurs, namely the declaration accompanying submitted work. Where institutional rules require a student to declare the assistance used, undisclosed machine generation becomes a false declaration,

¹⁵David Boud and Rebeca Soler, "Sustainable Assessment Revisited," *Assessment & Evaluation in Higher Education* 41, no. 3 (2016): 400–413.

¹⁶Dawson, *Defending Assessment Security in a Digital World*, 45–62.

¹⁷Regulation of the Minister of National Education Number 17 of 2010, Article 1(1).

¹⁸Law Number 12 of 2012, Article 70(1)–(2).

an offence that is evidentially tractable because it concerns what the student stated rather than how the text was produced.

The quality assurance framework supports this approach, since it locates responsibility for assessment design and academic standards with the programme and requires alignment between learning outcomes and assessment.¹⁹ An institution may therefore regulate disclosure and redesign assessment without further national instrument. What is required nationally is clarification that the plagiarism definition does not exhaust academic misconduct, so that institutions are not driven to strained interpretations of an instrument designed for a different problem.

D. Implications

Three implications follow for institutional policy. First, regulate disclosure rather than use. A rule requiring declaration of assistance is enforceable, equitable, and evidentially sound, and it preserves the information institutions need in order to understand actual practice. Second, design assessment for process evidence, accepting that this is costly and prioritising accordingly, with the highest-stakes assessments redesigned first. Third, treat artificial intelligence literacy, meaning the capacity to use such tools critically and to evaluate their output, as a graduate competence to be taught and assessed rather than as a risk to be suppressed; international guidance has moved in this direction, and the competence is now plainly required for professional practice.²⁰

A caution is warranted against the assumption that these measures resolve the underlying problem. They address the governance of conduct, not the deeper question of what competences remain worth certifying when routine textual production can be delegated. That question is properly curricular, and it will not be answered by integrity policy.

IV. CONCLUSION

This review asked how the discourse on generative artificial intelligence and academic integrity has developed, what the limits of the dominant institutional strategies are, and whether Indonesian regulation adequately reaches the conduct at issue. It found a clear trajectory from containment toward pedagogical redesign; that detection fails the requirement of evidential sufficiency and blanket prohibition fails enforceability and equity, while assessment redesign satisfies all three criteria at a cost that explains its slower adoption; and that the Indonesian plagiarism framework, constructed around appropriation of another party's work, does not straightforwardly reach undisclosed machine generation, so that the tractable locus of regulation is the declaration accompanying submitted work.

The theoretical implication is that academic integrity should be analysed as a problem of assessment design and evidential sufficiency rather than as a problem of student character, since the strategies that treat it as the latter perform worst on precisely

¹⁹Regulation of the Minister of Education, Culture, Research and Technology Number 53 of 2023 concerning Quality Assurance of Higher Education, Articles 5 and 17.

²⁰UNESCO, *Guidance for Generative AI in Education and Research*, 30–40.

the criteria that disciplinary decisions require. The practical recommendations are to regulate disclosure rather than use, to redesign high-stakes assessment for process evidence, to treat artificial intelligence literacy as a taught and assessed competence, and to clarify at the national level that the plagiarism definition does not exhaust academic misconduct. Given the pace of change in this field, subsequent research should examine the actual operation of disclosure regimes where they have been adopted, and should test empirically whether process-oriented assessment produces the integrity outcomes its proponents claim.

REFERENCES

- Boud, David, and Rebeca Soler. "Sustainable Assessment Revisited." *Assessment & Evaluation in Higher Education* 41, no. 3 (2016): 400–413.
- Bretag, Tracey, ed. *Handbook of Academic Integrity*. Singapore: Springer, 2016.
- Cotton, Debby R. E., Peter A. Cotton, and J. Reuben Shipway. "Chatting and Cheating: Ensuring Academic Integrity in the Era of ChatGPT." *Innovations in Education and Teaching International* 61, no. 2 (2024): 228–240.
- Dawson, Phillip. *Defending Assessment Security in a Digital World: Preventing E-Cheating and Supporting Academic Integrity in Higher Education*. London: Routledge, 2021.
- Fleckenstein, Johanna, Jennifer Meyer, Thorben Jansen, Stefan D. Keller, Olaf Köller, and Jens Möller. "Do Teachers Spot AI? Evaluating the Detectability of AI-Generated Texts among Student Essays." *Computers and Education: Artificial Intelligence* 6 (2024): 100209.
- Grant, Maria J., and Andrew Booth. "A Typology of Reviews: An Analysis of 14 Review Types and Associated Methodologies." *Health Information and Libraries Journal* 26, no. 2 (2009): 91–108.
- McCabe, Donald L., Kenneth D. Butterfield, and Linda K. Treviño. *Cheating in College: Why Students Do It and What Educators Can Do about It*. Baltimore: Johns Hopkins University Press, 2012.
- Rudolph, Jürgen, Samson Tan, and Shannon Tan. "ChatGPT: Bullshit Spewer or the End of Traditional Assessments in Higher Education?" *Journal of Applied Learning and Teaching* 6, no. 1 (2023): 342–363.
- Tlili, Ahmed, et al. "Mapping the Landscape of Generative AI in Education: A Systematic Review." *Computers and Education: Artificial Intelligence* 5 (2024): 100167.
- UNESCO. *Guidance for Generative AI in Education and Research*. Paris: UNESCO, 2023.
- Wu, Fan, Yang Dang, and Manli Li. "A Systematic Review of Responses, Attitudes, and Utilization Behaviors on Generative AI for Teaching and Learning in Higher Education." *Behavioral Sciences* 15, no. 4 (2025): 467. <https://doi.org/10.3390/bs15040467>.

Regulations and Policy Documents

- Law Number 12 of 2012 concerning Higher Education.
- Regulation of the Minister of National Education Number 17 of 2010 concerning the Prevention and Handling of Plagiarism in Higher Education.
- Regulation of the Minister of Education, Culture, Research and Technology Number 53 of 2023 concerning Quality Assurance of Higher Education.