INTERNATIONAL JOURNAL OF ACADEMIC RESEARCH IN BUSINESS & SOCIAL SCIENCES

Vol 16, Issue 5, (2026) E-ISSN: 2222-6990

Designing AI-Resistant Assessments in Higher Education: A Systematic Literature Review

Siti Norliana Ghazali, Jebakumari Selvarani Ebenezer, Nurul Bazilah Abd Hamid, Janaki Manokaran

Academy of Language Studies, Universiti Teknologi MARA, Malaysia

Email: jebakumari@uitm.edu.my, nbazilah@uitm.edu.my, janaki7819@uitm.edu.my

Corresponding Author Email: liana265@uitm.edu.my

DOI Link: http://dx.doi.org/10.6007/IJARBSS/v16-i5/28298

Published Date: 29 May 2026

Abstract

The rapid proliferation of generative artificial intelligence (AI), particularly ChatGPT, has disrupted conventional assessment practices in higher education. AI has enabled common tasks like reports and essays to be completed with minimal human authorship, raising pressing concerns about academic integrity, validity, and pedagogical relevance. This systematic literature review synthesizes recent empirical and conceptual papers to examine three interrelated questions: how AI challenges conventional assessment, which design strategies show promise in mitigating misuse, and where significant gaps remain in current research. Drawing on a PRISMA-guided review of 28 peer-reviewed studies published between 2022 and 2025, the analysis identifies five dominant themes; the structural limitations of traditional assessments, the shortcomings of AI-detection technologies, AI-resistant assessment design principles, the importance of authentic and process-oriented assessment and the need for establishing transparent institutional policies. Findings indicate that relying on ChatGPT detection technologies is insufficient, while authentic, process-based, and AI-integrated approaches are more practical and sustainable. The review argues that the rise of generative AI should not be viewed merely as a threat, but a catalyst for revising assessment designs in higher education.

Keywords: AI, ChatGPT, Assessments, Higher Education, Design

Introduction

The rapid advancement of generative artificial intelligence (AI) technologies has begun to reshape long-established assumptions about teaching, learning, and assessment in higher education. Large language models (LLMs) like ChatGPT, first released publicly in November, represents one of the most prominent and widely adopted examples of this new generation of AI tools, prompting widespread debate about their implications for academic integrity and assessment validity. AI enables students to generate coherent essays, synthesize various academic sources and produce reports and assignments with minimal human input, often at

a level of fluency that approximates or exceeds average student performance (Rudolph et al., 2023; Susnjak & McIntosh, 2024). Traditional tasks and assessments that had long been treated as reliable indicators of learning, now raise the questions of authorship and originality which cannot be taken for granted. This development signals that conventional assessments may no longer reliably measure student understanding, skill development, or independent reasoning when similar outputs can be generated with a few prompts by generative AI systems (Perkins et al, 2024).

Unlike earlier educational technologies such as calculators that can automate discrete computational steps and spell-checkers that can refine surface-level language, AI is not merely an assistive tool, but a system capable of simulating behaviours such as reasoning and synthesizing and producing outputs traditionally associated with higher-order human cognitive processes. Generative AI can autonomously produce coherent arguments, synthesize sources, and mimic specific writing styles, structures, and expectations that are typical of different academic fields or disciplines, from science papers to literary essays. This complicates efforts to distinguish between independent student work and AI-generated content (Perkins et al., 2024; Xia et al., 2024). When assessment outputs can be generated with minimal engagement from the learners, accepting such products like reports and essays as valid indicators of learning has become increasingly uncertain and difficult to determine (Patrick et al., 2025; Gamage et al., 2023).

Empirical evidence suggests that AI is perceived by students as a low-risk and efficient tools that can be utilized to complete academic tasks, particularly in tertiary level institutions that do not have a clear policy on the use of AI or where they are inconsistently enforced (Cotton et al., 2024; Güner et al., 2024). Ortiz-Bonnin and Blahopoulou (2025), find that students operating under unclear or inconsistent AI policies are significantly more likely to engage in dishonest use, when the universities fail to define what is permitted, restricted, or prohibited AI use. Students' behaviour is also shaped by whether they believe improper AI use can be detected and whether consequences are predictable and fair. Low perceived risk encourages experimentation with misconduct, not necessarily because students reject integrity norms, but because institutional signals suggest those norms are not meaningfully enforced (Luo, 2024; Li,2024) Normative expectations such as widespread, unchallenged ChatGPT use are also a factor is using ChatGPT in completing academic work (Leaton et al,2025).

In response to this perceived threat, many universities have initially turned to technological solutions, mainly deploying AI-detection tools. However, a growing body of empirical research has cast doubt on the reliability, fairness, and long-term viability of these approaches. Detection systems have been shown to generate both false positives where human-written works are misclassified as AI-generated and false negatives that fail to flag sophisticated AI-produced outputs (Perkins et al., 2024; Xia et al., 2024). Even when combined with expert human judgment, detection accuracy remains inconsistent across disciplines, genres, and linguistic contexts (Susnjak & McIntosh, 2024). Beyond technical limitations, detection-centred strategies raise ethical concerns related to transparency, cautioning that such tools may undermine institutional trust between students and educators and position assessment as a surveillance mechanism rather than a learning-oriented practice (Selwyn, 2022; Smolansky et al., 2023).

Studies suggest that the challenge posed by generative AI is not merely a problem of detection or enforcement, but a more fundamental issue of assessment design (Alkouk & Khlaif, 2024; Gundu, 2024). Traditional assessments like reports and essays that focus on the end products do not really provide valid evidence of learning such as original thinking and understanding concepts (Chaudhry et al., 2023; Rudolph et al., 2023). Patrick et al. (2025) note that the persistence of traditional, product-oriented assessment formats, particularly those that can be completed asynchronously and without supervision renders higher education vulnerable to AI misuse. Tasks commonly intended to measure higher-order cognitive skills such as analysis, evaluation, and synthesis are increasingly replicable by ChatGPT, raising questions about whether such assessments can still serve as meaningful indicators of student learning (Patrick et al., 2025; Gamage et al., 2023). As a result, the current situation has prompted calls for a paradigm shift towards more authentic, process-oriented assessments that are grounded in local or specific contexts (Peters & Angelov, 2025; Evangelista, 2024).

Besides making changes to traditional assessment, another emerging strand of research advocates for the integration of AI into assessment practices rather than its outright prohibition. From this perspective, developing students' ability to use AI critically, transparently, and ethically can be considered as part of the educational objectives (Dai et al., 2023; Bhullar et al., 2024). Such approaches require clear institutional policies that define acceptable uses of AI, promote transparency, and align assessment practices with evolving technological realities (Corbin et al., 2025; Moorhouse et al., 2023). However, studies indicate that integrating ChatGPT in assessment is a challenge as students and lecturers are unclear about acceptable ChatGPT use due to inconsistent policies across institutions, with many educators reporting uncertainty regarding expectations and enforcement (Smolansky et al., 2023; Corbin et al., 2025).

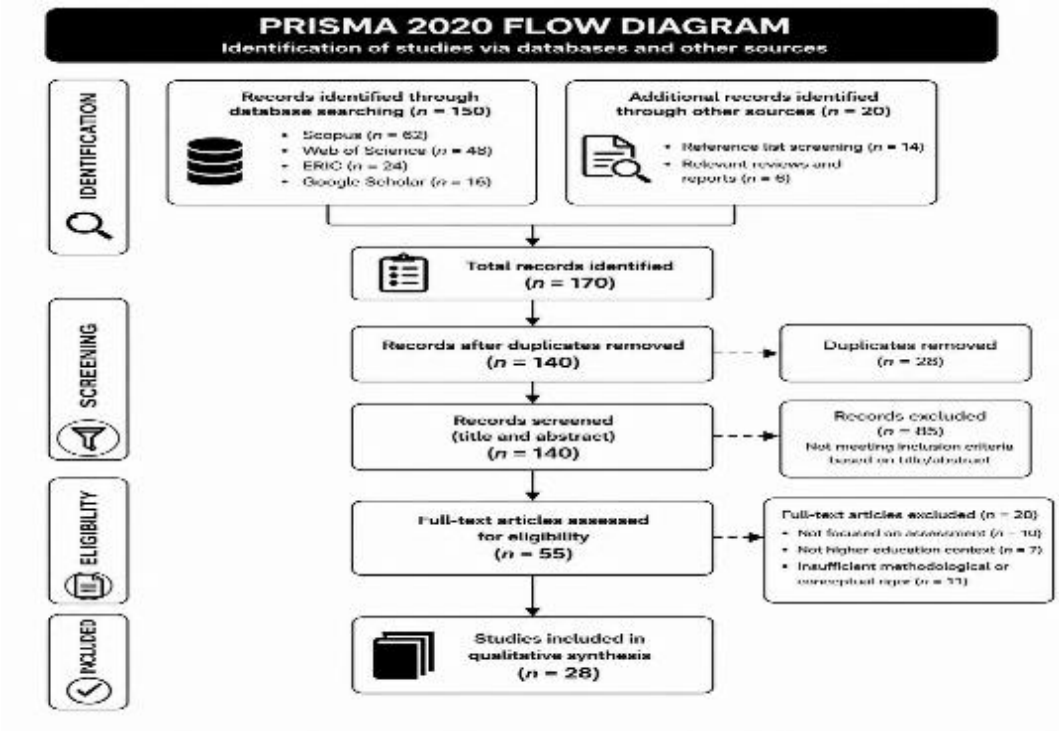
This current phenomenon of generative AI in higher education assessment is not just a technological issue, it also raises questions on how learning, authorship, and academic performance are evaluated. As AI evolves and can convincingly replicate higher-order thinking tasks like critical writing and reasoning, uncertainty emerges regarding the validity and integrity of traditional methods of assessment. While existing studies have discussed AI detection, academic integrity and assessment design, the literature remains fragmented and lacks a comprehensive synthesis on how assessment could be redesigned to cope with the current phenomenon. This creates a strong rationale for conducting the present systematic literature review, which seeks to consolidate evidence, identify emerging assessment design strategies and study the implications of AI-mediated learning environments for future assessment practices.

(1) What challenges does ChatGPT pose to traditional assessment practices in higher education?
(2) What strategies and design principles have been proposed to create AI-resistant assessments?
(3) What gaps exist in the current literature, and what future research directions are needed?

Methodology

This study adopts a systematic literature review (SLR) guided by the PRISMA 2020 framework (Page et al., 2021) to ensure transparency, rigor, and reproducibility. A structured search was conducted across Scopus, Web of Science, ERIC, and Google Scholar using combinations of keywords such as "ChatGPT," "generative AI," "artificial intelligence," "assessment," and "higher education." The search was limited to publications from 2022 to 2025 to capture recent developments following the emergence of ChatGPT in 2022. A total of 170 records were identified, with duplicates removed to yield 140 unique studies. Titles and abstracts were screened against predefined inclusion criteria which focus on higher education, relevance to generative AI and assessment, and publication in peer-reviewed or reputable sources, resulting in the exclusion of 85 studies.

The remaining 55 articles underwent full-text review, where studies were excluded if they did not directly address assessment (n = 10), focused on non-higher education contexts (n = 7), or lacked sufficient rigor (n = 8), leaving 28 studies for final inclusion. Data were analyzed using thematic synthesis, combining inductive and deductive coding to identify key patterns across the literature. This process resulted in five central themes: limitations of traditional assessment, limitations of AI detection approaches, principles for AI-resistant assessment design, authentic and process-oriented assessment practices, and transparency and institutional policy alignment. Consistent application of selection criteria and iterative cross-study comparison enhanced the reliability and depth of the synthesis.



Findings and Discussion

Limitations of Traditional Higher Education Assessment

A consistent and strongly evidenced finding across the reviewed literature is the growing inadequacy of traditional assessment formats in current AI-enhanced learning environment. Essay-based coursework, take-home assignments, and unsupervised online examinations are repeatedly identified as vulnerable because they rely on assumptions that students execute

their work independently and the completed tasks clearly show students' thinking processes, level of understanding and actual effort (Rudolph et al., 2023; Susnjak & McIntosh, 2024; Smolansky et al., 2023). However, the emergence of generative AI systems such as ChatGPT undermines these assumptions, as such tools can produce highly coherent, contextually appropriate, and academically styled responses that closely mimics human-generated text. This makes it increasingly difficult to determine how much understanding or work comes from the students (Rudolph et al., 2023; Susnjak & McIntosh, 2024; Chaudhry et al., 2023; Xia et al., 2024). As a result, strong performance in assessments no longer always reflects genuine learning or subject mastery.

From an assessment theory perspective, the emergence of ChatGPT presents a significant threat to construct validity, as students' scores in assessments may no longer accurately reflect the learning process that they are intended to measure (Chaudhry et al., 2023). Conventional assessments, particularly essays and reports have long been treated as indirect indicators of higher-order thinking such as analysing, evaluating, or creating ideas because they focus on visible features like formal academic language, clear logical structure and critical tone. However, these academic texts can now be reproduced convincingly by ChatGPT using suitable prompts with little human effort (Patrick et al., 2025; Rudolph et al., 2023). This creates a dilemma for examiners, who are increasingly unable to distinguish between authentic demonstrations of learning and outputs generated or heavily mediated by AI (Cotton et al., 2024; Smolansky et al., 2023). Moreover, AI is disrupting the way we traditionally understand levels of thinking in frameworks like Bloom's taxonomy where tasks such as writing essays or critiques have been assumed to require these higher-level cognitive processes. However, the evolving ChatGPT can now convincingly produce sophisticated, critical, and evaluative writing that looks like it involves deep thinking (Gamage et al., 2023; Xia et al., 2024). Consequently, the rapid integration of AI into academic work necessitates a fundamental rethinking of assessment design, moving towards approaches that foreground process, authenticity, and learner engagement to preserve validity and integrity in AI-mediated educational environments (Alkouk & Khlaif, 2024; Peters & Angelov, 2025).

However, the literature does not offer one single, unified diagnosis of the issue. Some studies perceive generative AI as a direct threat to the assessment of higher-order thinking, stressing that its ability to generate coherent, analytical, and evaluative responses enables students to bypass the cognitive processes that assessments are designed to reflect (Rudolph et al., 2023; Patrick et al., 2025; Smolansky et al., 2023). From this perspective, AI functions as an external disruptive force that weakens construct validity making assessment score no longer a reliable, trustworthy indicator of students' actual knowledge or skills as well as raising concerns about academic integrity (Cotton et al., 2024; Perkins et al., 2024; Susnjak & McIntosh, 2024). In contrast, other studies pinpoint the problem not in AI itself but in limitations of assessment design. Traditional assessments focus mainly on the final product such as essays, without showing how the students develop their ideas or reach their conclusions. As a result, they fail to capture key aspects of learning like reasoning, problem solving, collaboration and justification, making it harder to tell whether the work reflects genuine understanding (Dai et al., 2023; Noroozi et al., 2024; Gundu, 2024).

From this perspective, generative AI is not the main problem but rather a tool that reveals existing flaws in how assessments have been designed. This difference in views reflects a

broader divide; some researchers view AI as something that disrupts assessments and should be controlled while others see it as an opportunity to rethink and improve assessment by focusing more on how students learn instead of just what they submit (Peters & Angelov, 2025; Alkouk & Khlaif, 2024; Xia et al., 2024).

Across multiple studies, there is consistent evidence that students are increasingly integrating generative AI into their academic work in ways that challenge traditional assumptions about independent work and authorship (Cotton et al., 2024; Güner et al., 2024; Smolansky et al., 2023; Susnjak & McIntosh, 2024; Perkins et al., 2024; Corbin et al., 2025; Ortiz Bonnin & Blahopoulou, 2025; Xia et al., 2024). Cotton et al. (2024), argue that students increasingly perceive ChatGPT as a low-risk tool due to the difficulty of detection and the absence of clear institutional policies which contributes to the rapid normalisation of using AI to complete assessments, particularly where detection tools are weak or absent. Similarly, Güner et al. (2024) observe that students tend to make decisions about AI use based on what is convenient or useful such as when it helps them complete work faster rather than based on formal rules, academic integrity principles, or what is considered ethical or acceptable. Importantly, Ortiz Bonnin and Blahopoulou (2025) demonstrate that patterns of misuse strongly correlate with vague policy and unclear institutional regulation, rather than inherent student dishonesty. This aligns with broader findings across the literature that student behaviour is shaped by assessment design, perceived risk, and clarity of expectations, rather than simple intent to cheat (Smolansky et al., 2023; Corbin et al., 2025). Collectively, this evidence suggests that the rise in AI-assisted work is the result of misalignment between assessment practices, university policies, and evolving digital tools, highlighting more robust, AI-resilient assessment design.

The literature also identifies clear disciplinary variations in vulnerability to generative AI, with writing-intensive disciplines like humanities, social sciences, certain business studies and law being especially exposed (Bhullar et al., 2024; Lyanda et al., 2024). In these disciplines, common assessment tasks like literature reviews, reflective essays, and policy critiques can now be generated fluently by AI, making it hard to differentiate between student and AI-assisted work. By contrast, fields such as medicine, engineering, and the creative arts show greater resilience because they rely on observable practice and demonstrations, such as clinical simulations, laboratory work, design prototypes, or live performances (Lyanda et al., 2024; Noroozi et al., 2024). However, this resilience is increasingly fragile as AI tools are rapidly expanding into areas once considered less exposed, including code generating in programming assignments, design production and statistical analysis or modelling, thus extending vulnerability to a wider range of disciplines (Bhullar et al., 2024; Xia et al., 2024). The assumption that certain fields are shielded becomes increasingly unrealistic as AI abilities continue to evolve. Thus, these findings imply that traditional assessment formats are not merely at risk of misuse, but they no longer align with how knowledge is produced, accessed, and demonstrated in practice in the current AI-mediated higher learning environments (Dai et al., 2023; Gamage et al., 2023).

Limitations of AI Detection Approaches in University Contexts

Many universities have used AI-detection technologies as a preventative measure as a response to traditional assessment being increasingly vulnerable in AI-enriched environment, However, multiple independent studies, using different methods and datasets, arrive at

similar conclusion that AI-detection tools are unreliable, limited, and problematic. Findings suggest that AI detection tools are inconsistent and inaccurate, frequently producing both false positives by classifying human-written texts as AI-generated and false negatives**,** meaning they fail to detect AI use in the content. This undermining both reliability and fairness in scoring assessment (Perkins et al., 2024; Xia et al., 2024; Susnjak & McIntosh, 2024). For example, Perkins et al. (2024) prove that using different detection tools can give different results for the exact same piece of writing when analysing for AI-generated content. For example, one run with one detector might say a text is 80% likely to be AI-generated, while another check with a different tool might say 30% or even 0%, even though the text has not changed. Due to the inconsistency, these scores cannot be taken as proof. This is especially concerning in situations such as grading, academic misconduct cases, or disciplinary decision because unreliable results could result unfair judgments, such as wrongly accusing a student or failing to detect actual misuse particularly when texts are lightly edited or combined with human input (Susnjak and McIntosh (2024).

Besides technical limitations, scholars argue that using AI-detection tools is essentially a behaviourist and surveillance-oriented model of assessment, which focuses on monitoring, controlling, and identifying those who violate regulations rather than providing meaningful learning (Selwyn, 2022; Corbin et al., 2025). Thus, student assessment is treated as something to be regulated through detection software, and plagiarism checks. Selwyn (2022) criticizes the use of these detection tools, suggesting that assessments focus on the end products rather than the learning process. Similarly, Corbin et al. (2025) highlight that universities tend to focus on acceptable and unacceptable rather than engaging students in understanding authorship, responsibility and learning. In contrast, constructivist and authentic assessment approaches perceive learning as students actively engaging with tasks and contexts, rather than simply producing correct answers. Therefore, instead of trying to detect AI use in students' work, educators should design assessments that necessitates students to demonstrate their thinking processes, such as through drafts, reflections, or oral explanations, making it harder to rely on outputs generated by AI (Noroozi et al. ,2024). This contrast reveals a broader pedagogical tension between control-based approaches, which rely on monitoring and enforcement and design-based approaches, which seek to embed integrity within the assessments themselves (Alkouk & Khlaif, 2024; Peters & Angelov, 2025). While AI-detection strategies may offer short-term deterrence, they are increasingly viewed as inadequate and potentially counterproductive, as they can damage trust and fail to address weaknesses in assessment design (Smolansky et al., 2023; Cotton et al., 2024).

AI tools evolve rapidly and the idea of developing newer tools to detect its use in assignments can be futile as these detection systems become obsolete as AI continues to develop (Perkins et al., 2024; Gamage et al., 2023). Studies demonstrate that as models such as ChatGPT produce progressively coherent, contextually relevant and human-like responses, detection tools struggle to differentiate AI-generated and human-produced content, particularly when the texts are lightly amended by humans (Perkins et al., 2024; Susnjak & McIntosh, 2024). Perkins et al. (2024) show that detection systems can produce inconsistent and unstable scores, while Xia et al. (2024) report that no AI-detection tool can produce reliable accuracy across disciplines. In addition, Gamage et al. (2023) emphasizes that continuous updates to generative AIs rapidly outpace the development of detection tools, creating a persistent lag that limits their effectiveness over time. In contrast, Kooli and Yusuf (2025) provide a more

balanced view, accepting the limitations of detection tools while arguing that it can still play a supplementary role within a broader assessment strategy. For examples, detection tools could be used initially to screen and identify the amount of AI-generated content, followed by reviews by students to make appropriate amendments, rather than serving as definitive proof of misconduct. This aligns with suggestions from Perkins et al. (2024), who propose instead of relying solely on technology, it should be combined with expert judgments from faculty members to provide a more balanced and informed decision, though it still cannot guarantee complete fairness or accuracy. Overall, the literature agrees on the view that while detection can provide limited supporting role, it cannot serve as the only viable solution.

The use of detection tools is further complicated by ethical issues, particularly when the algorithms that these tools rely on to come up with their judgements are not transparent and are not made available to the public. Such tools raise serious concerns about who is responsible for decisions and whether fair procedures are followed, especially where students may be accused of plagiarism without clear evidence (Smolansky et al., 2023; Moorhouse et al., 2023; Perkins et al., 2024). For example, AI detection tools commonly provide scores without explaining how these are derived, making it hard for students to challenge the scores or for institutions to justify disciplinary actions. The risk of thi software incorrectly flagging student-written work as AI-generated content undermines principles of fairness and justice which are vital in valid assessment practices (Xia et al., 2024; Susnjak & McIntosh, 2024). Reliance on detection technologies also promotes the culture of surveillance in education, where students are monitored and evaluated through automated systems rather than trusted as active participants in their learning (Selwyn, 2022; Cotton et al., 2024). This can damage trust between students and institutions, which is essential to promote effective learning and academic integrity (Smolansky et al., 2023; Corbin et al., 2025). When learners sense they are being constantly monitored or judged by unreliable systems, they may become more strategic or defensive in their behaviour, rather than genuinely engaging in their learning process. Moreover, studies find that AI detection software tend to misclassify writing produced by non-native English speakers or students with diverse linguistic styles as AI-generated (Perkins et al., 2024).

In short, the reviewed literature suggests that detection-based approaches are not only technically insufficient but also theoretically and ethically misaligned with contemporary pedagogical goals. While they may offer short-term improvement, they do not address the structural vulnerabilities of assessment design, emphasizing the need for more sustainable alternatives.

Principles for ChatGPT-Resistant Assessment Design in Higher Education

In contrast to detection-centred strategies, the reviewed studies articulate principles for designing assessments that are more resistant to AI misuse. These principles are to redesign assessment to foreground reasoning, context, and learning processes which means shifting the focus of assessment from just the finished product to how learners think, learn, and arrive at their conclusions (Alkouk & Khlaif, 2024; Gundu, 2024; Peters & Angelov, 2025; Xia et al., 2024). AI resistant assessments focus on process-oriented tasks which are structured to show evidence of how students learn and think rather than what they produce (Dai et al., 2023; Noroozi et al., 2024). Studies indicate that tasks which require students to draw on local knowledge or context, personal experience, or course-specific materials and individual

project can substantially reduce the use of generic AI content (Alkouk & Khlaif, 2024; Gundu, 2024). This is because large language models like ChatGPT are less capable of generating accurate or meaningful responses when tasks depend on unique datasets, recent classroom interactions, or institution-specific scenarios (Alkouk & Khlaif, 2024; Gundu, 2024; Bhullar et al., 2024). For example, academicians report greater confidence in assessments that require students to analyse data they have personally collected, such as lab results, fieldwork observations as these cannot be easily imitated by AI. Similarly, tasks that require learners to reflect on in-class discussions, debates, or group activities require elements of specificity that exceed AI's general-purpose models (Noroozi et al., 2024; Lyanda et al., 2024). They also promote authenticity and relevance, aligning assessment more closely with real-world problem-solving (Peters & Angelov, 2025; Gamage et al., 2023).

Personalising assessments by including task that require students to discuss individual experiences, local contexts, or course-specific activities are acknowledged as more resistant to AI imitation because they demand knowledge not commonly provided by generic AI data (Alkouk & Khlaif, 2024; Gundu, 2024; Bhullar et al., 2024). However, scholars caution against assuming personalisation alone ensures AI proof assessment. Noroozi et al. (2024) find that AI can still generate realistic and believable "personal" responses, such as reflections or examples that *seems* to come from a student's own experience. Therefore, simply adding surface details like asking students to include personal experiences or local context is not enough. Instead, assessment should require students to explain their reasoning, justify their choices, or connect their experiences to course concepts that can be verified (Noroozi et al., 2024; Dai et al., 2023; Gamage et al., 2023). This changes the focus from what learners produce to how and why they arrive at their conclusions (Patrick et al., 2025; Peters & Angelov, 2025). Evangelista (2024) and Lyanda et al. (2024) highlight the value of iterative and scaffolded assessment designs, Iterative and scaffolded design where students complete their work in stages (from proposal to draft, feedback and revision) with lecturers giving support and guidance in each step. This allows instructors to trace tasks development over time and also reduce the incentive and opportunity for last-minute AI substitution, as students must engage continuously rather than produce a single polished output (Gundu, 2024; Alkouk & Khlaif, 2024).

The literature also underscores the importance of formal frameworks and assessment criteria in supporting AI-resistant design. Alvira et al. (2025) introduce and validate the *PANDORA rubric*, a systematic tool designed to help educators evaluate and redesign assessments in response to generative AI**.** It provides a systematic way to diagnose *why* an assignment is vulnerable and how it can be improved. It works by appraising the main elements of task such as the extent to which it includes context-specific information, justification of decisions, and visible learning processes, helping educators identify where assignments may rely too heavily on generic outputs that AI can easily replicate. Similarly, Fernández-Sánchez et al. (2024) show that rubrics created collaboratively by educators (and in some cases students), improve transparency and shared understanding of expectations, particularly regarding acceptable AI use. For example, their study highlights how clear rubrics can explicitly require students to record AI use, justify how texts are adapted, or explain why certain AI tools are used, making the whole learning process more transparent. Besides developing rubrics, the literature emphasises that these frameworks help move assessment away from implicit assumptions toward clearly articulated and transparent standards of practice and reduce ambiguity that

often leads to abuse of AI in assessment (Kooli & Yusuf, 2025; Ortiz Bonnin & Blahopoulou, 2025).

A particularly significant issue centres on whether AI should be excluded from assessment or explicitly integrated into it, reflecting fundamentally different perspectives. Some authors specifically define "AI-resistant" assessment as minimising or restricting AI use to preserve independent learning and academic integrity (Rudolph et al., 2023; Cotton et al., 2024; Smolansky et al., 2023). However, an alternative and increasingly influential position argues that total exclusion is both impractical and restrictive from a teaching perspective, especially since AI tools are commonly utilised in academic and professional settings. Researchers such as Dai et al. (2023) and Bhullar et al. (2024) recommend integrating AI use in assessments, where students are required to not only to use AI tools but also to critically evaluate, adapt, and justify their outputs. For example, tasks might ask students to generate an AI-produced response and then critique its accuracy, identify biases, improve its reasoning, or compare it with alternative approaches (Dai et al., 2023; Gamage et al., 2023). However, others are concerned about overreliance on AI, especially in areas like writing, critical thinking, and problem-solving (Selwyn, 2022; Perkins et al., 2024). For instance, Selwyn (2022) warns that excessive dependence on AI tools may lead students to abandon the process of learning and reducing opportunities to develop independent analytical skills. Similarly, Perkins et al. (2024) highlight the danger that students may accept AI-produced content indiscriminately, limiting their capabilities to assess evidence or detect errors. Studies caution that without explicit guidance, students may use AI in ways to save time over understanding (Güner et al., 2024; Ortiz Bonnin & Blahopoulou, 2025).

This debate reflects a broader shift among academicians, moving from "AI-resistant" to "AI-resilient" assessment paradigms (Peters & Angelov, 2025; Xia et al., 2024). Rather than attempting to block or avoid AI use entirely, AI-resilient approaches aim to design assessments that incorporate AI in ways that enhance, rather than replace, human thinking. For example, assessments may require students to disclose AI use, reflect on how it influences their work, and demonstrate independent reasoning alongside AI-supported work (Kooli & Yusuf, 2025; Fernández-Sánchez et al., 2024). In this way, the emphasis moves from controlling and detecting AI usage to developing students' capacity to use AI critically, ethically, and responsibly, aligning assessment with the realities of contemporary digital learning environment.

Authentic, Real-World, and Process-Oriented University Assessment

Principles of AI-resistant assessment design emphasize authentic, real-world tasks as the main strategy for maintaining academic integrity in AI-mediated contexts. Across the reviewed literature, authenticity is consistently identified as one of the most effective mechanisms for limiting AI substitution. Authentic assessments resemble professional practice but, more importantly, require contextual judgment, where students make decisions based on specific situations, and the integration of multiple knowledge sources, including theories, data, personal experience, and peer or instructor feedback (Gamage et al., 2023; Peters & Angelov, 2025; Xia et al., 2024). These tasks typically involve complex, applied activities such as case-based analyses, design projects, experiential learning, policy proposals, and prototype development (Peters & Angelov, 2025). Dai et al. (2023) further argue that embedding assessment within project-based learning environments shifts the emphasis

toward problem-solving processes rather than final products, thereby reducing the effectiveness of AI as a shortcut. The literature also highlights the importance of interactive and performative formats such as oral examinations, live presentations, and in-class problem-solving which require students to articulate reasoning in real time and respond to questions, making them difficult to use AI tools (Alkouk & Khlaif, 2024; Corbin et al., 2025). Hybrid designs that combine written work with presentations or demonstrations further strengthen the reliability and authenticity of assessment.

Complementing this focus on authenticity is an emphasis on process-oriented assessment, which emphasizes on making the student's learning process clearly observable to the instructor. Rather than evaluating only the final products, process-based approaches provide insight into how students construct and develop understanding across different stages of learning. Such tasks might include multi-stage submissions, reflective journals, and portfolio-based assessments (Evangelista, 2024; Lyanda et al., 2024; Xia et al., 2024). For example, following stages like drafting, followed by feedback and revision discourage last-minute AI substitution and provide traceable evidence of intellectual progression. Gundu (2024) similarly notes that continuous assessment reduces opportunities for outsourcing work to AI by embedding evaluation within an ongoing learning process rather than a single submission point. In addition, tasks that require students to explain their reasoning, justify decisions, and evaluate their own learning are less vulnerable to AI replication which struggle to produce authentic, contextually grounded reflection (Patrick et al., 2025; Fernández-Sánchez et al., 2024). In short, process-oriented strategies reduce AI use by but by providing longitudinal, multifaceted evidence of learning that prioritizes reasoning, engagement, and developmental growth over static performance outcomes.

Oral and interactive components are repeatedly highlighted as powerful complement to written assessment in AI-mediated contexts, particularly for validating authorship and demonstrating depth of understanding. Noroozi et al. (2024) demonstrate that collaborative argumentation tasks—where students must engage with, challenge, and respond to peers' opinions such as debates and forums create dynamic and unpredictable interactions that are difficult to emulate. Alkouk and Khlaif (2024) report findings from faculty workshop advocate the effectiveness of short viva-style discussions, presentations, and justification interviews that require students to explain and defend their reasoning in class. These are interrogative checkpoints but instead function as dialogic opportunities where students demonstrate authorship through explanation, clarification, and engagement with questions (Corbin et al., 2025; Gundu, 2024). Such interactions are immediate and unpredictable which make students difficult to rely on AI-generated products, especially when students are asked to extend or explain their ideas beyond prepared responses (Patrick et al., 2025; Perkins et al., 2024). Research further shows that interactive activities like conducting problem-solving tasks in class, group discussions, and live presentation also require dynamic interactions makes students rely less on AI (Noroozi et al., 2024; Xia et al., 2024). These approaches emphasise knowledge construction through dialogue and participation (Selwyn, 2022; Gamage et al., 2023) and when combined with written works, they offer more robust, multi-modal indication of student learning (Peters & Angelov, 2025; Alkouk & Khlaif, 2024).

However, studies also highlight significant challenges in implementing authentic, process-based assessment, particularly in terms of scalability, consistency, and resource demands.

Although these approaches are widely supported, they often require extensive redesign of curriculum, assessment criteria, and feedback practices, placing more work on both lecturers and students (Bhullar et al., 2024; Gundu, 2024; Evangelista, 2024). For example, scaffolded, multi-stage assessment submissions, portfolios, and project-based tasks demand continuous monitoring, iterative feedback, and more refined evaluation of student progress, which can be difficult for large classes. Lyanda et al. (2024) note that implementing these AI resilient assessment requires institutional and technological support. Xia et al. (2024) further emphasise that some disciplines like medicine and engineering rely on standardised exams to ensure fairness and consistency, so shifting to flexible, real-world assessments can make it harder to grade students equally and reliably. Incorporating oral or interactive components like debates, presentations, and in-class validations introduces more logistical complexities which include time constraints, staff training, consistent assessment standards (Alkouk & Khlaif, 2024; Corbin et al., 2025; Perkins et al., 2024). Findings also identify the issues of reliability and potential subjectivity when evaluating complex, authentic tasks, particularly in the absence of clear rubrics or shared criteria (Fernández-Sánchez et al., 2024; Kooli & Yusuf, 2025). Hence, these challenges highlight the importance to balance pedagogical innovation with practical considerations of scalability, equity, and institutional infrastructure.

Transparency, Acceptable Use, and Institutional Policy Alignment

The final theme emphasizes that assessment redesign must be accompanied by clear, coherent, and transparent institutional policies regarding AI use. Empirical evidence indicates that unclear regulations significantly increase the likelihood of AI misuse (Ortiz-Bonnin & Blahopoulou, 2025; Güner et al., 2024). However, while these studies highlight the behavioural impact of clear policy, Corbin et al. (2025) extend the discussion by arguing that policies must also define the pedagogical purpose of AI use by explaining the need to use it to enhance learning and the skills that can be developed, not merely the boundaries.

Institutional responses remain highly varied. Moorhouse et al. (2023) illustrate a shift toward AI-inclusive policies among leading universities such as University of Cambridge which permits limited use of AI that must be acknowledged and University of Oxford which allows AI use to generate ideas and language support but requires transparency and clearly prohibits submitting AI-generated work as student's own product. However, Bhullar et al. (2024) note that implementation remains inconsistent and often fragmented. Kooli and Yusuf (2025) also highlight governance challenges, particularly in relation to accountability and bias in AI-assisted processes. Universities must deal with complex issues about fairness and responsibility, especially when decisions are partly made by AI tools.

An important development in the literature is the growing support of transparency-based policy models, which prioritise clear disclosure over prohibition in AI application. Rather than rejecting AI use, several studies recommend compelling students to declare and justify how AI tools are utilized in their work, making transparency a core mechanism for maintaining academic integrity (Fernández-Sánchez et al., 2024; Corbin et al., 2025). This approach aligns with established principles of honesty and accountability while acknowledging the widespread integration of AI in universities (Bhullar et al., 2024; Moorhouse et al., 2023). Importantly, clear disclosure policies reduce ambiguity, which has been shown to contribute to misuse Ortiz-Bonnin & Blahopoulou, 2025). Fernández-Sánchez et al. (2024) demonstrate that embedding transparency requirements in assessment rubrics improves clarity and

supports consistent evaluation. Overall, transparency-based models represent a practical for navigating AI use in higher education.

Importantly, the findings also highlight the critical role of educators in implementing AI policies. Smolansky et al. (2023) report that when instructors lack confidence, training, or a shared understanding of AI guidelines, enforcement becomes inconsistent, leading to confusion among students and a potential loss of trust in assessment processes. In contrast, studies indicate that when institutions provide concrete support such as exemplary scenarios of acceptable and unacceptable AI use, discipline-specific guidelines, and professional development in AI, educators are better able to understand and apply policy consistently (Moorhouse et al., 2023; Corbin et al., 2025). For example, some universities have introduced annotated assessment samples showing how AI can be transparently integrated, while others offer training workshops to help staff design AI-aligned rubrics and tasks. Fernández-Sánchez et al. (2024) further demonstrate that embedding AI-use criteria within rubrics supports consistent evaluation and reduces ambiguity in marking. Similarly, Bhullar et al. (2024) emphasise that without institutional support, policy implementation will be difficult. These findings underscore that policy is not a static but a dynamic framework that is enacted shaped by educators and contextual application. In short, the literature suggests that effective policy must be closely aligned with teaching and assessment practices to ensure that expectations, practices, and learning goals are coherently aligned.

Research Gaps and Future Directions

Despite the increasing number of studies addressing generative AI and assessment, the reviewed literature reveals several persistent gaps that limit both theoretical consolidation and practical implementation. One of the most pressing gaps concerns many proposed assessment designs lack strong, consistent, and widely applicable empirical evidence as most have limited or preliminary findings. Although many studies advocate authentic, process-oriented, and AI-inclusion assessment models, they remain conceptual or based on small case studies, faculty workshops, or practitioner reflections (Alkouk & Khlaif, 2024; Dai et al., 2023; Noroozi et al., 2024). There is a notable absence of large, longitudinal, research examining how redesigned assessments perform across disciplines and institutional types. Without such evidence, claims regarding the effectiveness and scalability of AI-resistant assessment remain inconclusive.

A related gap concerns the limited development and application of measurement and validity frameworks in AI-mediated assessment. Although several studies identify construct validity as a core challenge, particularly the difficulty of ensuring that assessment outputs accurately reflect student learning rather than AI assistance (Patrick et al., 2025; Gamage et al., 2023; Rudolph et al., 2023) very few studies clearly define validity in measurable terms. Much of the existing work remains conceptual or design-oriented, with limited empirical investigation into how validity can be sutained or redefined in AI-integrated environments (Xia et al., 2024; Perkins et al., 2024). Moreover, there is a lack of explicit alignment between learning outcomes, assessment design, and the interpretive claims made about student performance, which is fundamental to validity theory. Future research would benefit from more rigorous and theory-informed approaches, including the use of established validation frameworks and comparative study designs. Research that examines traditional and redesigned assessments against shared learning outcomes could provide critical insights into how generative AI

impacts learning and whether new assessment formats really capture higher-order thinking and authentic learning (Noroozi et al., 2024; Gundu, 2024).

Finally, there is limited engagement with learner perspectives on redesigning assessment, representing a critical gap in understanding the effectiveness and sustainability of AI-augmented practices. While existing studies primarily focus on student use of AI tools—highlighting perceptions of ChatGPT as convenient, low-risk, and widely normalised for completing assignments (Cotton et al., 2024; Güner et al., 2024; Ortiz-Bonnin & Blahopoulou, 2025) few examine students experience using assessment models which are authentic, process-oriented, or AI-integrated. Student perceptions of fairness, workload, and legitimacy is important as they strongly influence engagement and compliance with academic integrity expectations (Smolansky et al., 2023; Bhullar et al., 2024). For example, authentic and process-oriented assessments may be perceived as more meaningful but also more demanding, particularly when they involve continuous submissions, reflection, or collaboration (Lyanda et al., 2024; Evangelista, 2024). Similarly, transparency-based AI policies that require disclosure may enhance clarity but could also raise concerns about surveillance or unequal access to AI tools across students (Fernández-Sánchez et al., 2024; Moorhouse et al., 2023). Without systematic investigation into how students interpret and respond to these evolving practices, reforms risk misalignment between pedagogical intent and learner experience.

Conclusion

The emergence of generative AI systems particularly ChatGPT has brought into sharp focus the limitations of long-standing assessment practices in higher education. This review shows that traditional, product-oriented formats such as essays, reports, and unsupervised take-home examinations are increasingly unable to serve as reliable indicators of student learning in AI-assisted environments. While detection-based approaches have been widely adopted as short-term responses, they remain technically unreliable and are often misaligned with contemporary pedagogical principles that prioritise trust, learning, and student development.

A clear consensus across the literature is that sustainable responses to generative AI lie not in tighter surveillance, but in the redesign of assessment itself. Authentic, process-oriented, and specific, contextually grounded tasks, supported by iterative development, reflective justification, and transparent criteria offer a better foundation for maintaining both validity and integrity. Rather than excluding AI, these approaches reposition assessment to evaluate forms of understanding that are less easily replicated by automated systems, including reasoning, judgment, and the ability to integrate and apply knowledge in context.

At the same time, effective assessment reform depends on sustained alignment between policy, pedagogy, and practice, rather than treating these domains as discrete or sequential interventions. Clear institutional guidance on acceptable AI usemust be meaningfully embedded within assessment design, teaching strategies, and feedback practices. Equally important is the development of students' AI literacy, not only in technical terms but in relation to critical evaluation, ethical judgment, and responsible integration of AI into learning processes. When such alignment is achieved, expectations become transparent and coherent, supporting both student confidence and instructor consistency

Overall, this review contributes by synthesising a fragmented and rapidly evolving body of research into a coherent, theoretically informed perspective that connects assessment design, empirical evidence, and institutional policy. By integrating insights across disciplines and methodological approaches, it highlights both convergences and tensions within the literature, offering a more unified understanding of emerging practices. Crucially, the review reframes generative AI not simply as a problem to be controlled, but as a catalyst for rethinking the fundamental purposes of assessment. In doing so, it underscores a shift away from inherited, product-focused models toward more meaningful, process-driven, and context-sensitive approaches.

References

Alkouk, W. A., & Khlaif, Z. N. (2024). AI-resistant assessments in higher education: Practical insights from faculty training workshops. *Frontiers in Education, 9*, Article 1499495. https://doi.org/10.3389/feduc.2024.1499495

Alvira, N. B., Bannister, P., & Santamaría Urbieta, A. (2025). Validating the PANDORA GenAI susceptibility rubric for higher education assessment: A field test of translation and interpreting BA assignments. *Higher Education Quarterly*, Article e70056. https://doi.org/10.1111/hequ.70056

Bhullar, P. S., Joshi, M., & Chugh, R. (2024). ChatGPT in higher education: A synthesis of the literature and a future research agenda. *Education and Information Technologies, 29*, 21501–21522. https://doi.org/10.1007/s10639-024-12723-x

Chaudhry, I. S., Sarwary, S. A. M., El Refae, G. A., & Chabchoub, H. (2023). Time to revisit existing student performance evaluation approaches in higher education in the era of ChatGPT: A case study. *Cogent Education, 10*(1), Article 2210461. https://doi.org/10.1080/2331186X.2023.2210461

Corbin, T., Dawson, P., Nicola-Richmond, K., & Partridge, H. (2025). 'Where's the line? It's an absurd line': Towards a framework for acceptable uses of AI in assessment. *Assessment & Evaluation in Higher Education, 50*(5), 705–717. https://doi.org/10.1080/02602938.2025.2456207

Cotton, D. R. E., Cotton, P. A., & Shipway, J. R. (2024). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International, 61*(2), 228–239. https://doi.org/10.1080/14703297.2023.2190148

Dai, Y., Liu, A., & Lim, C. P. (2023). Reconceptualizing ChatGPT and generative AI as a student-driven innovation in higher education. *Procedia CIRP, 119*, 84–90. https://doi.org/10.1016/j.procir.2023.05.002

Evangelista, E. D. L. (2024). Ensuring academic integrity in the age of ChatGPT: Rethinking exam design, assessment strategies, and ethical AI policies in higher education. *Contemporary Educational Technology, 17*(1), Article ep559. https://doi.org/10.30935/cedtech/15775

Fernández-Sánchez, A., Lorenzo-Castiñeiras, J. J., & Sánchez-Bello, A. (2024). Navigating the future of pedagogy: The integration of AI tools in developing educational assessment rubrics. *European Journal of Education*. Advance online publication. https://doi.org/10.1111/ejed.12826

Gamage, K. A. A., Dehideniya, S. C. P., Xu, Z., & Tang, X. (2023). ChatGPT and higher education assessments: More opportunities than concerns? *Journal of Applied Learning and Teaching, 6*(2), 358–369. https://doi.org/10.37074/jalt.2023.6.2.32

Gundu, T. (2024). Strategies for e-assessments in the era of generative artificial intelligence. *European Journal of E-Learning, 22*(7). https://doi.org/10.34190/ejel.22.7.3477

Güner, H., Er, E., Akçapınar, G., & Khalil, M. (2024). *From chalkboards to AI-powered learning: Students' attitudes and perspectives on use of ChatGPT in educational settings*. Educational Technology & Society, 27(2), 386–404. https://doi.org/10.30191/ETS.202404_27(2).TP05

Kooli, C., & Yusuf, N. (2024).*Transforming educational assessment: Insights into the use of ChatGPT and large language models in grading*. *International Journal of Human–Computer Interaction*. https://doi.org/10.1080/10447318.2024.2338330

Leaton Gray, S., Edsall, D., & Parapadakis, D. (2025). AI-based digital cheating at university and the case for new ethical pedagogies. *Journal of Academic Ethics, 23*, 2069–2086. https://doi.org/10.1007/s10805-025-09642-y

Li, Z. (2024). *Generative AI in higher education academic assignments: Policy implications from a systematic review of student and teacher perceptions* (Master's thesis). Massachusetts Institute of Technology. https://hdl.handle.net/1721.1/155977

Luo, J. (2024). A critical review of GenAI policies in higher education assessment: A call to reconsider the "originality" of students' work. *Assessment & Evaluation in Higher Education, 49*(5), 651–664. https://doi.org/10.1080/02602938.2024.2309963

Lyanda, J. N., Owidi, S. O., & Simiyu, A. M. (2024). *Rethinking higher education teaching and assessment in line with AI innovations: A systematic review and meta-analysis*. *African Journal of Empirical Research, 5*(3), 325–335. https://doi.org/10.51867/ajernet.5.3.30

Moorhouse, B. L., Yeo, M. A., & Wan, Y. (2023). Generative AI tools and assessment: Guidelines of the world's top-ranking universities. *Computers and Education Open, 5*, Article 100151. https://doi.org/10.1016/j.caeo.2023.100151

Noroozi, O., Soleimani, S., Farrokhnia, M., & Banihashem, S. K. (2024). Generative AI in education: Pedagogical, theoretical, and methodological perspectives. *International Journal of Technology in Education, 7*, 373–385. https://doi.org/10.46328/ijte.845

Ortiz-Bonnin, S., & Blahopoulou, J. (2025). Chat or cheat? Academic dishonesty, risk perceptions, and ChatGPT usage in higher education students. *Social Psychology of Education, 28*, Article 113. https://doi.org/10.1007/s11218-025-10080-2

Patrick, P. M., Yip, S. Y., & Campbell, C. (2025). Artificial intelligence and higher-order thinking: A systematic review of educator and student experiences and perspectives in higher education. *Higher Education Quarterly, 79*(4), Article e70069. https://doi.org/10.1111/hequ.70069

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ, 372*, n71. https://doi.org/10.1136/bmj.n71

Perkins, M., Roe, J., Postma, D. (2024). Detection of GPT-4 generated text in higher education: Combining academic judgement and software to identify generative AI tool misuse. *Journal of Academic Ethics, 22*, 89–113. https://doi.org/10.1007/s10805-023-09492-6

Peters, M., & Angelov, D. (2025). Redefining assessment tasks to promote students' creativity and integrity in the age of generative artificial intelligence. *International Journal for Educational Integrity, 21*, Article 25. https://doi.org/10.1007/s40979-025-00201-x

Rudolph, J., Tan, S., & Tan, S. (2023). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education? *Journal of Applied Learning and Teaching, 6*(1), 342–363. https://doi.org/10.37074/jalt.2023.6.1.9

Selwyn, N. (2022). The future of AI and education: Some cautionary notes. *European Journal of Education, 57*, 620–631. https://doi.org/10.1111/ejed.12532

Smolansky, A., Cram, A., Raduescu, C., Zeivots, S., Huber, E., & Kizilcec, R. F. (2023). Educator and student perspectives on the impact of generative AI on assessments in higher education. In *Proceedings of the 10th ACM Conference on Learning at Scale*. https://doi.org/10.1145/3573051.3596191

Susnjak, T., & McIntosh, T. R. (2024). ChatGPT: The end of online exam integrity? *Education Sciences, 14*(6), Article 656. https://doi.org/10.3390/educsci14060656

Xia, Q., Weng, X., Ouyang, F., et al. (2024). A scoping review on how generative artificial intelligence transforms assessment in higher education. *International Journal of Educational Technology in Higher Education, 21*, Article 40. https://doi.org/10.1186/s41239-024-00468-z