Journal of Innovative Research (JIR)

ISSN: 2837-6706 (ONLINE)

VOLUME 4 ISSUE 3 (2026)

PUBLISHED BY
E-PALIU PUBLISHERS, DELAWARE, USA

Journal of Innovative Research (JIR)

Volume 4 Issue 3, Year 2026
ISSN: 2837-6706 (Online)
DOI: https://doi.org/10.54536/jir.v4i3.9008
https://journals.e-palli.com/home/index.php/jir

Bounded AI Assistance in Higher Education: Re-Aligning Pedagogy, Assessment and Institutional Governance

Richard Mulenga[1*], Marien Matafwali[1]

Article Information

Received: August 22, 2026

Accepted: November 03, 2026

Published: November 21, 2026

Keywords

Bounded AI Assistance, Higher Education, Institutional AI Policy, Large Language Models, Pedagogical Reforms

ABSTRACT

The rise of large language models (LLMs) has led universities and other higher education institutions to rethink the ways in which knowledge is conveyed, learning is facilitated, and academic performance is evaluated. While such technologies present possibilities for tailored explanations, feedback that supports learning as it unfolds, and joint inquiry, they also give rise to concerns regarding inaccurate information, reliance on machines for thinking, academic integrity, and inequities in access. Drawing on a narrative synthesis of selected literature and institutional guidance, this conceptual paper explores the pedagogical ramifications of LLMs. The paper contends that integrating them effectively demands learning and research environments that bring together subject-specific knowledge, critical questioning, openness about bounded AI use, and demonstrable competence. Constructive alignment, self-regulated learning, cognitive load theory, and cognitive engagement supply the conceptual bases for linking intended learning outcomes, teaching and learning activities, and assessment. A cyclical process is put forward, comprising independent engagement, AI-assisted interaction under guidance, verification, application, and reflection. The paper goes on to elaborate dimensions of institutional policy, encompassing governance, curriculum, assessment, research integrity, privacy, intellectual property, equity, procurement, and quality assurance. Specific consideration is devoted to contexts with constrained ICT resources. The paper's conclusion is that the educational value of LLMs rests on purposeful design of specific institutional AI policy that is clear, transparent and progressive.

INTRODUCTION

Large language models (LLMs) are transforming the circumstances in which instruction and learning take place in the education sector in general and higher education in particular. The ability of LLMs to produce explanations, condense texts, compose arguments and write computer code opens up novel avenues for academic support (katyara *et al.*, 2025). At the same time, such capabilities pose a challenge to assessment methods and bounded Artificial Intelligence (AI) use that depend heavily on written outputs as proof of individual comprehension. Kasneci *et al.* (2023) point to possibilities for personalised learning and the creation of educational content, while stressing critical thinking, verification and human oversight. This is usually referred to as the "human in the loop, (HITL).".

The key pedagogical issue concerns how colleges and universities can safeguard substantive intellectual growth at a time when students are able to generate seemingly credible answers almost immediately. A well polished essay might be the product of genuine learning, of considerable AI assistance, or of a combination of the two. Therefore, higher education institutions must consequently rethink or re-align how submitted work relates to the competence that work is meant to demonstrate. This realignment is meant to elucidate questions of misconduct and connect it to the fundamental aims of higher education, that include, among others, disciplinary knowledge, autonomous judgement, ethical accountability and the capacity to apply learning to unfamiliar problems. UNESCO situates human agency, inclusion and educational appropriateness at the core of integrating generative AI (Miao, *et al.*, 2024). Existing frameworks address university AI governance, define permissible levels of AI use in assessment, and examine which cognitive tasks students can delegate as their competence develops (Chan, 2023; Perkins *et al.*, 2024; Salameh, 2026). However, a recent scoping review identifies a need for closer alignment between intended learning outcomes and assessment methods, stronger evidence of how students learn with AI, clearer reporting of governance practices, and greater attention to equity (Lin *et al.*, 2026). In African higher education, the empirical literature on generative AI and assessment redesign remains limited and is concentrated in a small number of countries, particularly Nigeria and South Africa (Antwi-Boampong *et al.*, 2026). These findings point to a design question: How can universities translate institutional AI policy into task-specific limits on LLM assistance while enabling students to demonstrate disciplinary competence under unequal conditions of access? This conceptual paper addresses that question by proposing a five-stage frame work of independent engagement, guided AI use, verification, application, and reflection. The cycle links the assistance permitted at each stage to intended learning outcomes, evidence of student competence, and institutional responsibilities for disclosure, equitable access, and quality assurance. Drawing on a selection of peer-reviewed literature and credible institutional policy documents, this article offers a conceptual discourse regarding the realignment

[1] Department of Economics, ZCAS University, Lusaka, Zambia
[*] Corresponding author's e-mail: richardmulenga2@gmail.com

of pedagogical approaches in the backdrop of the ever evolving and disruptive AI assisted in academics and research. It constitutes a narrative synthesis rather than a systematic review, and it presents no original empirical findings. The institutional AI bounded assistance dimensions and applications put forward in this study are intended for adaptation and evaluation, not empirically validated models nor assertions of universal effectiveness. The literature review brings to the fore what appear to be unresolved questions concerning how bounded AI assistance influences retention and transfer, how assessment can ascertain both disciplinary competence and responsible AI use, and which institutional arrangements sustain equitable implementation across settings that differ in resources. This article contributes to the literature by linking these gaps through a coherent pedagogical cycle framework and a discussion of institutional policy, offering them as a conceptual synthesis rather than as an established empirical framework.

Historical Context and Scope of the Evidence

The use of AI in higher education long predates today's scholarly discourses. Zawacki-Richter *et.al* (2019) catalogued uses spanning prediction, assessment, adaptive systems, and intelligent tutoring, while noting that links to pedagogical theory remained weak and that educational risks received limited critical scrutiny. This background is significant, since deploying a technology successfully and improving education integrity do not always achieve the same results. The adoption of LLMs ought to be assessed in the light of evolving questions concerning how students learn, how educators exercise judgement, and whose interests institutional systems ultimately serve. Lo's (2023) literature review documented the early educational literature that emerged after ChatGPT's release, noting possible teaching support as well as inaccuracies and challenges for assessment. However, given that this review's observation window was relatively early, it is valuable for grasping initial responses rather than for assessing what all later GPT models can do. Bond *et al.* (2024), surveyed reviews of AI in higher education more broadly. The authors urge greater attention to ethics, methodology, and context. Taken together, these studies point to the fact that AI assisted models and approaches warrant a careful synthesis in order to differentiate technological promise from demonstrated educational value. Accordingly, this article evaluates AI assisted approaches in four categories: Learning theory, empirical studies of educational outcomes, reviews of the research field, and normative or institutional guidance.

LITERATURE REVIEW

Large Language Models and the Changing Meaning of Learning

Large language models (LLMs) such as, among others, GPT, Claude, Gemini, DeepSeek, Quen and Mistral, produce responses based on patterns acquired during training together with information provided within the interaction. Although their fluency may render them valuable educational resources, persuasive phrasing does not amount to factual correctness or sound reasoning. To assess the responses they obtain, students need disciplinary knowledge. Possible uses encompass explanations, practice activities and feedback, alongside risks of bias, misleading content and overreliance on AI tools (Kasneci *et al.*, 2023). A key distinction separates successful completion of a task from lasting learning. Generating a strong answer with technological assistance does not necessarily imply that a student can explain that answer, reproduce the reasoning, or transfer the underlying knowledge to a novel situation. Bastani *et al.* (2025), in a study of high-school mathematics, found that unrestricted generative AI assistance could enhance performance during supported practice while weakening later performance once the assistance was withdrawn. These adverse effects were substantially reduced by a tutor built with learning safeguards. Given that the study focused on secondary education sector, its results should inform higher education with caution rather than be regarded as direct evidence applicable across university or college disciplines.

Evidence likewise shows the promise of carefully designed AI support. In a randomised controlled trial (RCT) with college physics students, Kestin *et al.* (2025) found that a purpose-built AI tutor yielded stronger learning outcomes than the active-learning classes used for comparison. This lends support to structured tutoring in that context; it does not demonstrate that unrestricted chatbot use is universally superior to classroom teaching.

Students Experience and the Limits of Perceived Usefulness

Chan and Hu (2023) conducted a survey of university students in Hong Kong, recording perceptions of generative AI that were largely positive while also noting apprehensions regarding its use. Findings of this study contend the inclusion of students in discussions about implementation. However, favourable perceptions ought not to be treated as equivalent to independent mastery. An AI tool may appear convenient precisely because it eliminates difficulty, even in cases where that difficulty constitutes an integral component of the intended learning activity. Evaluations, therefore, ought to differentiate among immediate assistance, understanding that is retained, and knowledge transfer. Thus, institutions ought to identify which outcome serves as the motivation for a proposed use and to gather evidence suited to that purpose, instead of merging the outcomes into a single assertion of effectiveness.

Constructive Alignment as a Foundation for Reform and Re-alignment

Constructive alignment denotes the deliberate interconnection of intended learning outcomes, the learning activities by means of which students build understanding, and the assessment tasks employed to assess whether the intended outcome has occurred (Biggs, 1996). When translated into the integration of LLMs in higher education and research, this principle obligates lecturers to take as their starting point the competence that

students are expected to develop. If the intended outcome is independent interpretation, then the assessment ought to make clear the students' own processes of reasoning. If the outcome concerns professional engagement with AI, the assessment should gauge how professionals steer, check and refine work that is supported by AI. An economics course, for instance, the lecturer might ask students to differentiate correlation from causation. Asking students to supply definitions offers only weak evidence of such competence, given that definitions can be produced automatically. A task that is better aligned requires students to critically evaluate an empirical claim, detect potential confounding factors, and justify a defensible interpretation. Pedagogy supported by LLMs should consequently bring together direct instruction, independent practice and guided inquiry. The relative weighting of these elements ought to differ according to discipline, stage of study and intended learning outcome.

Self Regulated Learning and Metacognitive Responsibility

Zimmerman (2002) conceptualizes self-regulated learning as an active process in which learners take charge of managing their own learning. When applied to the use of LLMs, this viewpoint draws attention to setting goals, monitoring comprehension, and evaluating progress. Before turning to AI assistance, students ought to determine what they aim to learn and then judge whether a given response enhances their own explanation. This application constitutes a theoretical interpretation rather than evidence that interacting with a chatbot as was the case Zimmerman (2002) study, automatically generates self-regulation. Lecturers may require students to identify an accepted suggestion, a rejected suggestion, and the evidence underlying each decision. Reflection ought to be proportionate and verified against actual performance.

Cognitive Load and the Design of Assistance

Sweller (1988) accounts for the ways in which the cognitive burden imposed by problem solving may influence the formation of knowledge structures. This provides grounds for differentiating genuinely useful instructional support from the mere reduction of difficulty across all tasks. A beginner confronting a complicated econometric problem might profit from a worked-out explanation of a single step yet could be overwhelmed by an extended response filled with unfamiliar terminology. Instructors ought to determine both the quantity and the ordering of assistance in accordance with what learners already know. One consequence for instruction supported by LLMs is that hints, worked examples, and staged questions ought to serve deliberate purposes. A model could clarify how a coefficient is to be interpreted prior to students' examination of the model's assumptions. Learners are then able to practise using a different example while progressively receiving diminished support.

Cognitive Engagement and Collaborative Knowledge Construction

A taxonomy proposed by Chi and Wylie (2014) differentiates between passive, active, constructive and interactive forms of engagement. This framework directs attention toward the cognitive operations learners perform, rather than toward outward signs of busyness. Reproducing an answer produced by a model, marking passages of it, and articulating the reasons for its incompleteness each represent a distinct mode of engagement. Across studies of undergraduate Science Technology Engineering and Mathematics (STEM) education, Freeman *et al.* (2014) reported advantages associated with active learning. Such findings lend support to deliberate student involvement, yet they do not demonstrate that activities built on LLMs are effective in particular.

In condensing the literature review, the preceding review of the literature shows that the pedagogical value of large language models (LLMs) rests on how they are designed and evaluated for teaching, and specifically on the contrast between support for completing tasks and support for learning independently. Constructive alignment ties together intended learning outcomes, learning activities and assessment (Biggs, 1996), whereas the cognitive engagement perspective focuses on how students explain, critique and apply information instead of simply accessing it (Chi & Wylie, 2014). From these perspectives it follows that content generated by AI ought to underpin purposeful intellectual activity. Evidence from empirical studies underscores the significance of instructional design. Kestin *et al.* (2025) reported learning gains associated with a purpose-built AI tutor in college physics, while Bastani *et al.* (2025) reported that unrestricted assistance enhanced performance during mathematics practice but reduced later performance without assistance among high-school students. Direct comparisons and generalisation are constrained by differences in settings and interventions. Together with feedback research that stresses students' active interpretation and use of comments (Nicol & Macfarlane-Dick, 2006), these findings provide a basis for bounded AI assistance. The literature additionally ties pedagogical reform to credible assessment and to accountable governance at the institutional level. Limitations in the reliability of AI-text detection, together with documented misclassification of writing by non-native English speakers, argue for caution in treating detector scores as decisive proof of misconduct or plagiarism (Liang *et al.*, 2023; Weber-Wulff *et al.*, 2023). Assessment should consequently pair explicit permissions for assistance and requirements for disclosure with credible demonstrations of individual competence. At the institutional level, Chan (2023) connects pedagogical, governance and operational responsibilities, and UNESCO stresses human agency, privacy, inclusion and educational suitability (Miao & Holmes, 2023). These perspectives indicate that innovation in the classroom depends on staff development, accessible tools and consistent implementation, even though published policies do not in themselves demonstrate educational effectiveness. Bond *et al.* (2024) underline the need for stronger ethical, methodological and contextual foundations in AI research within higher education. The resulting synthesis of the literature highlights what appear

to be unresolved questions: How bounded AI assistance affects retention and transfer, how assessment can establish both disciplinary competence and responsible use of AI, and which institutional arrangements enable equitable implementation across settings that differ in resources. This article contributes by linking these concerns through a coherent pedagogical cycle and a discussion of institutional policy, offering them as a conceptual synthesis to be tested rather than as an established framework

MATERIALS AND METHODS

This article adopts a narrative Synthesis Methodology, informed by Popay *et al.* (2006), to integrate heterogeneous evidence on large language models (LLMs), pedagogy, assessment and institutional governance in higher education. Narrative synthesis is an approach to integrating findings from multiple sources through structured comparison, interpretation and explanation. It is particularly useful when studies differ in their methods, populations, interventions or outcomes. In this paper, Narrative Synthesis approach offers a way to connect learning theories, evaluate empirical; studies and institutional policy guidance while preserving distinctions between these sources. Narrative synthesis is an approach/ design is different from a narrative literature review. The review design determines *how sources* are identified and selected, whereas the synthesis approach *determines how their findings are integrated (Popay, 2006;* Miao, & Cukurova,2024). An interpretivist philosophical orientation guides contextual interpretation (Boell & Cecez-Kecmanovic, 2014). Constructive alignment, self-regulated learning and cognitive load theory provide sensitising concepts. The synthesis asks: How can higher education institutions redesign pedagogy, assessment and governance to support meaningful learning and responsible LLM use? The publication window is 1 January 2020–31 December 2025, with searches conducted between January and August 2026. Earlier foundational theoretical and methodological sources are considered separately. The search locations include Scopus, Web of Science, ERIC, Google Scholar and relevant institutional repositories. Search terms used combine "large language models," "generative AI" and "ChatGPT" with "higher education," "pedagogy," "assessment," "academic integrity" and "institutional policy." Eligible sources comprised relevant empirical studies, theoretical contributions, evidence syntheses and institutional policy documents. Duplicates, promotional material and publications without a clear higher education connection were excluded.

RESULTS AND DISCUSSION

To conserve space, the results of the narrative synthesis are reported in Appendix as a populated evidence matrix. Twenty nine documents constituted the final sample that was analysed employing the narrative synthesis methodology. Nine pre-2020 literature sources were treated as foundational or background literature. Six key themes arising from the analysis are reported in Table 1:

Table 1: Key Themes from the Narrative Synthesis

Theme	Summary	Implication for Higher Education	Supporting Citations
1. AI assistance and independent learning	Designed tutoring can improve immediate learning; unrestricted assistance may weaken subsequent independent performance.	Combine guided assistance with opportunities to demonstrate competence independently.	Bastani *et al.* (2025); Kestin *et al.* (2025)
2. Established pedagogical principles	Alignment, cognitive engagement, self-regulation and cognitive load inform purposeful instructional design.	Align AI activities with learning outcomes, prior knowledge and assessment requirements.	Biggs (1996); Chi and Wylie (2014); Sweller (1988); Zimmerman (2002)
3. Perceived usefulness versus demonstrated learning	Satisfaction and convenience do not establish retained understanding or transferable competence.	Evaluate learning outcomes alongside student perceptions and adoption.	Bond *et al.* (2024); Chan and Hu (2023); Kasneci *et al.* (2023)
4. Academic integrity and assessment reform	Fabricated citations, unreliable detection and biased classifications challenge conventional integrity practices.	Require source verification, clear assistance rules and assessments revealing individual reasoning.	Liang *et al.* (2023); Walters and Wilder (2023); Weber-Wulff *et al.* (2023)
5. Institutional policy and AI literacy	Responsible integration requires coordinated governance, infrastructure and staff and student capabilities.	Develop institutional policies alongside training and practical implementation support.	Chan (2023); Miao and Cukurova (2024); Miao *et al.* (2024)
6. Human agency, equity and continuing evaluation	Human oversight, privacy, inclusion and ongoing risk assessment underpin responsible implementation.	Provide equitable access, assign accountability and monitor educational benefits and risks.	Miao and Holmes (2023); Tabassi (2023); UNESCO (2022)

Note: The themes integrate empirical findings, theoretical contributions and policy document analyses.

The six key themes outlined in Table 1 can be summarized as follows.

• **AI Assistance and Independent Learning:** Purpose-built AI tutoring can improve immediate learning, whereas unrestricted assistance may improve supported performance while weakening subsequent independent achievement (Bastani *et al.*, 2025; Kestin *et al.*, 2025).

• **Established Pedagogical Principles:** Constructive alignment, cognitive engagement, self-regulation and cognitive load provide foundations for designing purposeful AI-supported learning activities (Biggs, 1996; Chi & Wylie, 2014; Sweller, 1988; Zimmerman, 2002).

• **Perceived Usefulness versus Demonstrated Learning:** Positive student perceptions and technological convenience should be distinguished from evidence of retained knowledge, independent understanding and transferable competence (Bond *et al.*, 2024; Chan & Hu, 2023; Kasneci *et al.*, 2023).

• **Academic Integrity and Assessment Reform:** Fabricated references, unreliable AI-text detection and biased classifications reinforce the need for source verification, transparent assistance rules and assessments that reveal individual reasoning (Liang *et al.*, 2023; Walters & Wilder, 2023; Weber-Wulff *et al.*, 2023).

• **Institutional Policy and AI literacy:** Responsible integration requires coordinated pedagogical, governance and operational policies, supported by staff development and student AI competencies (Chan, 2023; Miao & Cukurova, 2024; Miao *et al.*, 2024).

• **Human Agency, Equity and Continuing Evaluation:** AI implementation should preserve human oversight, protect privacy, support equitable access and undergo continuing evaluation of educational benefits and risks (Miao & Holmes, 2023; Tabassi, 2023; UNESCO, 2022).

Rethinking Teaching and Learning Approaches

Inquiry and Problem Based Learning

Large language models (LLMs) may be integrated into inquiry-based learning tasks wherein learners explore questions, contrast competing explanations, and weigh the available evidence. Students of public finance, for instance, might examine whether a hypothetical district ought to prioritise clinics, road maintenance, or cash transfers. Although a model could produce tentative arguments, students would assess budget constraints, distributional consequences, and implementation risks by drawing on credible evidence (Chan and Hu (2023). Students in business could appraise a generated strategy for market entry in light of customer interviews and industry data. Students of education could evaluate a generated lesson plan with respect to accuracy, accessibility, and suitability for the classroom. These constitute proposed instructional uses.

Guided Dialogue and Formative Support

Large language models (LLMs) can facilitate learning via explanations, practice questions, and feedback, on the condition that such applications are subjected to critical scrutiny (Kasneci *et al.*, 2023). Instructors are able to design interactions in which students initially tackle a problem, articulate their reasoning, and subsequently seek a hint or feedback concerning a particular difficulty. It is expected that students will revise their work and provide justification for why the revision is warranted. Formative usage ought to assist students in detecting learning gaps, rather than simply yielding a corrected answer.

Formative feedback rests upon a considerable educational evidence base. Black and Wiliam (1998) link enhanced formative assessment to learning, whereas Nicol and Macfarlane-Dick (2006) foreground the active part students play in producing and applying feedback. As applied in this context, the educational question concerns whether students comprehend the feedback and respond to it. Generating a greater volume of comments automatically constitutes an insufficient indicator of improvement. A lecturer may direct students to weigh AI feedback against a marking rubric and peer comments, choose a revision that is justified, and account for its impact.

Critical and Disciplinary AI literacy

Long and Magerko (2020) define AI literacy in terms of the competencies required to interact with AI and to evaluate it critically. This lends support to teaching students when assistance is appropriate and how to identify limitations. UNESCO's student competency framework encompasses human-centred, ethical, technical and design dimensions (Miao *et al.*, 2024). The concept of AI literacy goes well beyond the ability to craft effective prompts. Students must be able to check references, identify claims that lack support, and consider how bias or absent context may influence the answers produced. In an evidence-audit exercise, students can be asked to sort generated claims into categories such as supported, unsupported, misleading or context-dependent, to find credible sources, and to explain the corrections that are required. UNESCO's student competency framework encompasses human-centred, ethical, technical and design dimensions (Miao *et al.*, 2024). Since that framework is directed at school curricula, applying it in higher education calls for adaptation to disciplinary depth, adult learners and professional expectations. Walters and Wilder (2023) reported fabricated references and substantive bibliographic errors in the outputs of the model versions examined in their study. Their findings should not be regarded as current error rates for all systems. The pedagogical implication is that a reference which appears plausible must be independently verified for its existence, its bibliographic details and its relevance to the claim. Independent writing, calculation, reading and explanation remain necessary for intellectual development. This recommendation follows from the distinction between supported performance and

independent learning illustrated by Bastani *et al.* (2025) but requires evaluation within university settings.

Rethinking Assessment and Academic Integrity

Pedagogical Rationale of Bounded AI Assistance

Bounded AI assistance refers to the deliberate integration of artificial intelligence into teaching, learning and assessment within clearly defined educational boundaries (Miao & Holmes, 2023). These boundaries specify when students may use AI, the forms and extent of assistance permitted, the intellectual tasks they must undertake independently, and the evidence required to demonstrate learning. For example, students may use AI to obtain conceptual explanations, receive formative feedback or explore alternative perspectives, while retaining responsibility for developing arguments, verifying sources, interpreting findings and defending conclusions (Bond *et al.*, 2024), alongside evidence that carefully designed AI tutoring can support learning in particular higher education settings (Kestin *et al.*, 2025). An integration that places humans at the centre calls for careful consideration of who is able to reach AI-supported learning and who stands to gain from it. UNESCO (2022) foregrounds inclusion, equity, privacy and suitability for education. In Zambia and comparable settings with limited resources, implementation ought to start from local evidence regarding access to devices, connectivity, electricity and institutional support, instead of presuming that availability is uniform.

Equity concerns the quality of participation, not merely whether a device can be accessed. Students may vary in confidence, language proficiency and prior disciplinary knowledge, and these differences influence how effectively they can judge generated content. Chan and Hu (2023) demonstrate why student experience ought to shape implementation, whereas UNESCO (2022) situates fairness and human dignity within wider AI governance. Institutions ought to seek the views of affected groups and to analyse how benefits and burdens are distributed.

Assignments should embed local relevance by drawing on credible regional scholarship and contextual evidence. Sweller's (1988) cognitive load theory provides a framework for understanding how support, when sequenced appropriately, may facilitate the acquisition of knowledge. Nicol and Macfarlane-Dick (2006) highlight the importance of feedback that enhances students' ability to monitor and regulate their own learning. Extending these theories to LLMs indicates that assistance should target a particular learning challenge while preserving opportunities for students to reason, practise, and assess their progress. Table 2 illustrates the application of bounded AI assistance to the learning/lecture environments.

Boundaries require both clear instructions and practical implementation. A statement prohibiting complete answers does not ensure that a general-purpose chatbot will comply. Lecturers should therefore evaluate the

Table 2: Approaches to Bounded AI Assistance

Approach	How the boundary Operates	Illustrative Application
Attempt before assistance	Students complete an initial attempt before accessing AI support.	An econometrics student interprets a regression coefficient before requesting feedback on the interpretation.
Hints and guided questioning	AI provides prompts or incremental hints while students develop the solution.	A public finance student receives questions about distributional effects before evaluating a proposed tax.
Feedback on student work	AI comments on a student-generated draft; the student decides and explains the revisions.	A student evaluates feedback on a policy brief against the marking rubric and supporting evidence.
Assistance limited to particular tasks	The lecturer permits specified activities while reserving assessed intellectual work for the student.	Language editing is permitted, while developing the argument and interpreting findings remain the student's responsibility.
Assistance at designated stages	AI access is permitted during selected phases of an activity.	Students formulate a research question independently, use AI to explore alternatives, and independently defend their final choice.
Support that gradually decreases	Guidance is reduced as students demonstrate greater competence.	Students move from a worked example to partial hints and finally to independent problem solving.
Verification and justified revision	Students must check generated claims and justify their use.	A student verifies a suggested reference and explains whether it supports the relevant claim.
Independent demonstration after assistance	Students subsequently explain or apply the learning without AI.	Following assisted practice, students analyse a different dataset or answer a short -supervised question.

Source: Adapted from UNESCO's human-centred approach to generative AI in education (Miao & Holmes, 2023) and Lyu (2026)

responses students receive and design assessment around demonstrable understanding.

Process and Achievement

Assessment ought to furnish credible proof that students have attained the intended outcomes. The Tertiary Education Quality and Standards Agency (TEQSA, 2023) points to generative AI as a justification for revisiting assessment design, while still building on established assessment expertise. A well-balanced programme is able to pair supervised work that demonstrates individual competence with assignments in which any permitted AI assistance is explicitly evaluated. Evidence of process may comprise a research question, annotated sources, preliminary analysis, a draft, and an account of the revisions made. Where necessary, chosen AI interactions may be incorporated, once personal or confidential information has been removed. These materials shed light on how the work developed, yet they do not on their own establish authorship.

Boud and Falchikov (2006) contend that assessment ought to equip students for learning that continues beyond graduation. Within professional settings mediated by AI, this encompasses appraising the quality of information and identifying when a tool ought not to be depended upon. A programme is able to map assessment across its courses in order to determine where essential competencies are demonstrated under supervised conditions and where students practise responsible assistance. Mapping of this kind prevents all responsibility from being placed on a single assignment.

Authentic Tasks and Individual Explanation

Authentic assessment may take the form of a policy brief grounded in specified data, a case from a local organisation, or field observations. Contextual detail on its own does not avert outsourcing, given that LLMs are able to respond to detailed scenarios. Such work ought to be paired with opportunities for students to explain decisions, respond to questions, and apply concepts to an altered problem. A short oral discussion might inquire why a particular method was selected, what evidence would change the conclusion, and which limitation is most significant(UNESCO, 2022). Planning at the programme level is able to allocate these activities among courses in a way that avoids both excessive duplication and an unreasonable workload.

Assistance that is Permitted, and Procedures that are Fair

Instructions for assessment ought to specify whether the use of AI is prohibited, limited, permitted or required; to identify which activities are allowed; and to clarify what is expected in terms of disclosure and verification. One course might allow language editing while still demanding independent analysis, whereas a different course might require generated material to serve as an object of critique. When rules are clear at the task level, expectations become open to review. Weber-Wulff *et al.* (2023) identified considerable reliability issues in the AI-text detectors they examined, among them incorrect classifications and diminished effectiveness once text had been modified. Their findings pertain to the tools and testing conditions that were examined in 2023, and not to every detector developed afterwards. Even so, they offer a basis for not treating detector scores as adequate proof of misconduct or plagiarism.

Education on integrity and the Fairness of Enforcement

Cotton *et al.* (2024) explore the opportunities and the academic integrity challenges that arise in connection with ChatGPT. An institutional response ought to make acceptable collaboration and assistance intelligible by means of worked examples. Students are able to discuss cases that involve grammar support, idea generation, code completion and generated analysis, and then to compare their own judgements with the assignment rules as they actually stand. Activities of this kind help to link integrity expectations with the ordinary decisions students make in their studies.

Liang *et al.* (2023) found that the detectors examined in their study frequently misclassified writing produced by non-native speakers of English. This lends further weight to the fairness concerns that Weber-Wulff *et al.* (2023) raised. The findings are specific to their samples and tools, yet they caution against assuming that statistical features of writing are enough to establish authorship. In multilingual institutions, students ought to be given a clear opportunity to respond to concerns, and legitimate language support should be kept distinct from substantive assistance that is prohibited. Fair procedures also call for consistent communication before assessment takes place. In Appendix 2, the implementation matrix translates the proposed five-stage cycle for bounded AI assistance into practical guidance for university policy, course design, and quality assurance. It clarifies which decisions belong to university leaders and programme teams and what instructors and students are expected to do at each stage. It also identifies the evidence universities can review to assess implementation and refine their guidance.

The Evolving Role of Educators in Higher Education Institutions

The rise and evolution of LLMs heightens the significance of professional judgement when choosing activities, foreseeing misconceptions, and determining whether students are building competence. Miao and Cukurova (2024) delineate five dimensions of AI competence for teachers. A human-centred mindset, AI ethics, foundations and applications, AI pedagogy, and AI for professional learning. These dimensions can be converted by institutions into discipline-specific workshops and collaborative redesign efforts. One valuable result is a revised assignment that specifies learning outcomes, guidelines on permitted use, assessment criteria, and a

plan for evaluation. Time and recognition are required for lecturers to carry out this work. Responsibilities can be shared, and reusable support materials developed, by academic departments, libraries, teaching centres, and information technology units.

Rather than simply receiving institutional systems after procurement, academic staff should take part in evaluating them. Zawacki-Richter *et al.* (2019) stress the value of educators' perspectives, while Bond *et al.* (2024) underscore interdisciplinary collaboration. Distinguishing a technically impressive demonstration from an activity that genuinely supports the curriculum requires teaching expertise. Complementary knowledge is contributed by librarians, accessibility specialists, and students. Implementation can be aided by collegial moderation. The effort involved in redesign, moderation, and student support should be acknowledged by institutions. Assertions that AI inevitably saves staff time should be examined once verification, correction, training, and maintenance have been considered.

Equity, Contextual and Institutional Responsibility

An integration that places humans at the centre calls for careful consideration of who is able to reach AI-supported learning and who stands to gain from it. UNESCO foregrounds inclusion, equity, privacy and suitability for education (Miao & Holmes, 2023; UNESCO 2022). When AI use is called for, students require a practicable way to reach an approved tool or a learning pathway of equivalent value. Consideration may be given to campus facilities and to access at scheduled times. Assessment ought not to reward access to paid services unintentionally where such access is unevenly distributed. Students should likewise examine whether generated explanations are suited to local institutions, languages and socioeconomic circumstances, drawing on pertinent Zambian evidence and professional practice. Equity concerns the quality of participation, not merely whether a device can be accessed. Students may vary in confidence, language proficiency and prior disciplinary knowledge, and these differences influence how effectively they can judge generated content. Chan and Hu (2023) demonstrate why student experience ought to shape implementation, whereas UNESCO (2022) situates fairness and human dignity within wider AI governance. Institutions ought to seek the views of affected groups and to analyse how benefits and burdens are distributed.

Institutional AI Development in Universities and Colleges

Effective pedagogical reform depends on institutional regulations and support mechanisms that reach beyond the boundaries of individual classrooms. UNESCO advocates integrated, human-centred approaches to the governance and assessment of generative AI (Miao & Holmes, 2023). An institutional AI policy should specify its scope, allocate decision-making authority, and assign responsibilities across teaching, research and administration, while providing practical guidance that remains adaptable as tools and practices evolve. Chan (2023) introduces an AI education policy framework encompassing pedagogical, governance and operational dimensions, drawing on evidence from students and staff in Hong Kong universities. Its principal contribution lies in linking educational aims to institutional responsibilities and organisational capacity UNESCO's Recommendation on the Ethics of Artificial Intelligence, adopted in 2021 and published in 2022, offers a broader ethical foundation for human oversight, fairness and responsible governance (UNESCO, 2022). Tabassi's (2023) voluntary AI Risk Management Framework provides a complementary organisational perspective. These documents serve distinct purposes: Ethical principles articulate values, whereas institutional procedures specify decision-making authority and the processes for evaluating and addressing risks.

Lessons from Institutional Practice

The University of Oxford's published summative assessment policy obliges assessment setters to state whether and in what manner AI may be used, ensure that assessment design corresponds to authorised assistance, define declaration requirements, and guarantee equal baseline access when such use is permitted (University of Oxford, 2025). The University of Cape Town announced a framework endorsed in June 2025 that structures engagement around AI literacy, assessment integrity, and educational innovation. Its development was informed by consultation with groups from across the institution (University of Cape Town, 2025). The reader should note that these cases are presented for illustration rather than as a comprehensive comparison or an assessment of relative effectiveness.

Governance and Accountability

Colleges and universities ought to designate an accountable executive sponsor and establish a suitable academic approval pathway. A multidisciplinary committee may comprise academic staff, students, research ethics personnel, librarians, information technology specialists, and staff with responsibility for accessibility and procurement. Responsibilities should be delineated explicitly: Academic bodies should establish educational standards, whereas designated operational units should evaluate systems and facilitate implementation. In a smaller college, these responsibilities may be integrated into existing committees. The policy should specify the systems and activities within its scope, including institutionally procured services, external tools employed for institutional work, and AI functions incorporated into existing software. It should differentiate advisory assistance from decisions with significant consequences and specify which uses necessitate further approval. These arrangements are proposed governance models, not accounts of requirements already adopted by any identified institution. A policy architecture may consist of

a stable institutional policy, operational standards subject to more frequent revision, and teaching guidance tailored to particular disciplines. For instance, the policy may mandate meaningful human oversight; an operational standard may establish approval conditions for sensitive data; and a course handbook may define authorised assistance. (Tabassi, 2023; University of Oxford, 2025). Any adaptation should remain firmly grounded in educational purposes. A risk framework informs institutional judgement without determining whether a learning activity has pedagogical value.

Curriculum and Pedagogical Standards

Policy should mandate that courses articulate how the use of AI contributes to their intended learning outcomes. Departments should determine which competencies students must develop independently and which professional AI capabilities graduates are expected to possess. The policy should preserve disciplinary judgement while defining baseline requirements for accuracy, inclusion and transparency. Constructive alignment offers a foundation for examining the links among learning outcomes, learning activities and assessment (Biggs, 1996; University of Cape Town, 2025).

Assessment and Academic Integrity

An institutional baseline should specify unambiguous assessment instructions, disclosure requirements and equitable procedures for handling misconduct. Departments may tailor these requirements to the characteristics of particular tasks. Where applicable, declaration forms should specify the tool used, the purpose of the assistance and the portions of the work it influenced. Disclosure cannot render otherwise prohibited assistance permissible, while permitted assistance does not absolve students of responsibility for accuracy Oxford's approach demonstrates the relationship between permissions defined for specific tasks and assessment design (University of Oxford, 2025).

Research Ethics and Scholarly Integrity

Research provisions should encompass AI-assisted literature reviews, analysis, coding, translation and writing. Researchers should authenticate references and outputs, record substantive uses and satisfy applicable journal, funder and institutional requirements. AI-generated data must not be presented as observations gathered from participants or from the real world. Human researchers continue to bear responsibility for interpretations and reported findings. Research ethics procedures should evaluate proposed uses of participant information in AI systems, including issues of consent, confidentiality and approved data management. Before unpublished manuscripts, confidential peer-review documents, or restricted datasets are transferred to any external system, the requisite authorisation must be obtained. These provisions operationalise the broader human-centred and privacy principles examined by Miao and Holmes (2023) as proposed institutional practices. Research supervision should safeguard the student's control over both the research problem and its interpretation. Supervisors may inquire into the effects of AI assistance on a coding decision or literature search, but such explanations should be assessed together with research records and disciplinary competence. Research training should explicitly address fabricated citations and differentiate exploratory suggestions generated by AI from verified scholarly evidence (Walters & Wilder, 2023; University of Cape Town, 2025).

Data Privacy and Information Security

Institutions should categorise the data permissible for entry into each system and establish criteria for selecting approved tools. Such reviews should address retention practices, access controls, deletion procedures, providers' use of submitted information, and incident-response mechanisms. Students and employees require clear examples of sensitive information, such as marks, identifiers, personnel records, and confidential research materials. Privacy and security should be assessed before adoption, rather than relying exclusively on individual users' review of service terms. (Miao and Cukurova, 2024; University of Oxford, 2025).

Intellectual Property and Attribution

Policy should establish clear expectations concerning the use of copyrighted readings, teaching materials, student work, and unpublished research. It should differentiate acknowledgement of AI assistance from citation of the underlying scholarly evidence. References generated by AI must be verified against the original publications. Questions concerning ownership and licensing should be resolved by reference to applicable institutional policies and relevant agreements, rather than through a general assumption that all outputs are unrestricted (Walters & Wilder, 2023, UNESCO, 2022).

Equity Accessibility and Cultural Relevance

Institutions should determine whether required tools are accessible to students with disabilities and financially affordable for all student groups. Comparable access arrangements should be established before compulsory use is introduced. Teaching should incorporate materials, languages and examples that are locally relevant, while generated content should be reviewed for cultural and contextual gaps. Students should be provided with an explicit procedure for reporting access problems without incurring academic penalties (UNESCO, 2022).

Staff Development and Student Support

AI policy should allocate responsibility and funding for induction, ongoing professional development and support tailored to specific disciplines (Marigocio, 2026). Training should address critical evaluation, data management, assessment design and disclosure practices. Libraries may assist with source verification and information literacy,

whereas teaching centres may contribute expertise in learning design. The competency dimensions proposed by Miao and Cukurova (2024) offer a valuable framework for organising educator development. Students should be involved in consultations concerning disclosure requirements, access arrangements and support resources. Chan (2023) assigns students a role in policy development, while Chan and Hu (2023) illustrate the value of attending to students' experiences.

Procurement and Sustainable Resourcing

Before acquiring services, institutions should assess their educational suitability, accessibility, security, overall cost and reliance on any single supplier. Pilot initiatives should be governed by explicit criteria for continuation and termination. Budgets should account for training, support and connectivity in addition to subscription charges. Resource and environmental consequences may be taken into account when dependable information is available. These constitute proposed procurement criteria rather than evidence that any specific commercial system meets them (University of Oxford, 2025).

Administrative Use and Human Oversight

AI can assist with routine drafting, information retrieval and service enquiries, but decisions with significant consequences require review by an accountable human. Policy should specify boundaries for applications influencing admission, grading, financial assistance and employment (Bond *et al.* 2024). Individuals should understand the procedures available for requesting correction or contesting an unfavourable decision. Administrative efficiency should be considered together with accuracy, fairness, and the potential effects of mistakes.

Quality Assurance and Policy Review

The policy should identify the person responsible for implementation, establish a review date, and set out a procedure for reporting problems. Monitoring may address independent learning, retention after a delay, disparities in access, the quality of assessment, workload, and costs. Changes to tools and reported incidents may warrant bringing the review forward. Adoption levels and user satisfaction, considered in isolation, do not adequately demonstrate educational value. A feasible monitoring framework may focus on a limited group of indicators, including independent performance on shared tasks, performance after a delay, the quality of verification, reported access problems, staff workload, and incidents that have been resolved. Institutions should establish baseline measurements before implementation begins. Bond *et al.* (2024) advocate greater contextual and methodological rigour in research, whereas Tabassi (2023) provides a foundation for ongoing organisational evaluation.

Table 3: A Proposed Pedagogical Cyclical Framework for LLM Integration

Stage	Main learning activity	Evidence of learning
Independent engagement	Attempt the task using existing knowledge and prescribed resources.	Initial explanation, calculation or problem formulation.
Guided AI interaction	Use permitted assistance for hints, questions, feedback or alternative perspectives.	Selected interactions and reasons for seeking assistance.
Verification and critique	Check claims, methods and references against credible evidence.	Corrections, source evaluation and justified acceptance or rejection.
Application and defence	Apply the learning to a new problem and explain key decisions.	Revised analysis, practical demonstration or oral explanation.
Reflection	Identify what was learned, what remains uncertain and how assistance affected the process.	A concise account of learning and remaining difficulties.

Source: Authors' Elaboration from Literature Review

Building upon the foregoing discussion, this article advances a cyclical framework comprising five stages as shown in Table 3.

The cycle or framework may be modified in accordance with class size, disciplinary context, and available resources.

Implementation and Future Research

Pedagogical innovations should be introduced by institutions through pilot programs that remain feasible to manage. Before broadening implementation, departments can identify suitable courses, define the anticipated educational benefits, provide lecturers with appropriate support, and compare the resulting outcomes. Evaluation should extend beyond measures of satisfaction and the quality of submitted work. The divergent findings reported by Bastani *et al.* (2025) and Kestin *et al.* (2025) underscore the importance of investigating specific learning designs instead of making generalized assertions about AI effectiveness. Additional research is particularly necessary in African higher education contexts, across diverse disciplines, and under varying resource conditions.

Methodological Limitations

This study draws on a narrative synthesis of selected scholarship and institutional guidance. Its coverage therefore depends on how sources were identified and selected, and relevant studies or policies may have been missed. The included sources also serve different purposes. Empirical studies report outcomes in specific

settings, learning theories offer explanations, and institutional policies prescribe practices. Differences in disciplines, student populations, LLM tools, and outcome measures limit direct comparisons across studies and the conclusions that can be drawn about the effects of AI assistance on learning. The proposed five-stage cycle and its accompanying governance recommendations are conceptual and have yet to be evaluated in practice. The study therefore cannot establish whether the cycle improves students' ability to apply disciplinary knowledge independently, strengthens assessment validity, or reduces inequalities in access.These limitations are particularly relevant to resource-constrained universities, where implementation requires local evaluation of whether students can participate equitably and demonstrate the same learning outcomes.

CONCLUSION

The rise and rapid evolution of large language models (LLMs) heighten the significance of professional judgement when choosing activities, foreseeing misconceptions, and determining whether students are building competence. Professional development ought to encompass curriculum design, assessment, feedback, and ethics in addition to technical operation. LLMs offer higher education institutions a significant opportunity to reexamine the ways in which they foster and assess intellectual development. Their educational value is determined by the interplay among learning outcomes, pedagogical design, student engagement, and assessment practices. An effective response integrates foundational knowledge with inquiry, structured support, verification, and independent demonstration of learning. It additionally necessitates explicit expectations regarding academic integrity, professional development for lecturers, and equitable access. Universities should assess their effectiveness according to the extent to which students cultivate understanding, sound judgement, and competencies applicable across contexts. These outcomes continue to constitute the core responsibilities of higher education within an educational environment increasingly influenced by generative AI. The study limitations are that narrative synthesis approach adopted in this study offers an interpretive reading of bounded AI assistance, it does not empirically validate the proposed model. Future studies should compare bounded AI assistance, unrestricted use and conventional instruction through experimental or quasi-experimental designs, measuring independent achievement, delayed retention and transfer. Longitudinal and mixed-methods research should examine differences across disciplines, student groups and resource settings. Institutional studies should also evaluate whether policy clarity, lecturer development, AI literacy and equitable access improve teaching, assessment credibility and student learning.

The empirical findings discussed in this paper come from prior studies of specific uses of large language models (LLMs). Their conclusions apply to the learners, tasks, tools, and outcomes examined in those studies. In particular, better performance while using an LLM and the ability to apply the same knowledge independently are distinct outcomes and require separate empirical evidence. The five-stage cycle of independent engagement, guided AI use, verification, application, and reflection is this paper's conceptual contribution. It draws on learning theory, empirical literature, and institutional guidance to specify when LLM assistance may be used and what students should demonstrate independently. The cycle and its accompanying governance recommendations require empirical evaluation to determine whether they improve independent learning, strengthen assessment validity, or support equitable participation in institutions with limited digital resources.

Acknowledgement

The authors thank the reviewers for their constructive comments on an earlier version of this manuscript. Their feedback helped clarify the proposed framework for bounded AI assistance and strengthen the discussion of its implications for teaching, assessment, and institutional governance.

Author Contributions

Dr. Richard Mulenga conceptualised the study, designed the narrative synthesis, and prepared the initial manuscript. Dr. Marien Matafwali contributed to the interpretation of the literature and critically revised the manuscript, with particular attention to the implications of bounded AI assistance for teaching and assessment. Both authors contributed to refining the proposed framework and reviewed and approved the final manuscript.

Declaration of Interests

The authors declare that they have no competing financial or non-financial interests relevant to this manuscript.

REFERENCES

Antwi-Boampong, A., Nketia, M. O., Loglo, F. S., Frimpong, N., & Osei-Fosuaa, L. (2026). Rethinking assessment in the age of generative AI: A systematic literature review of African higher education. *Social Sciences & Humanities Open, 14*, Article 103332. https://doi.org/10.1016/j.ssaho.2026.103332

Bastani, H., Bastani, O., Sungu, A., Ge, H., Kabakcı, Ö., & Mariman, R. (2025). Generative AI without guardrails can harm learning: Evidence from high school mathematics. *Proceedings of the National Academy of Sciences, 122*(26), Article e2422633122. https://doi.org/10.1073/pnas.2422633122

Biggs, J. (1996). Enhancing teaching through constructive alignment. *Higher Education, 32*(3), 347–364. https://doi.org/10.1007/BF00138871

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice, 5*(1), 7–74. https://doi.

org/10.1080/0969595980050102

Bond, M., Khosravi, H., De Laat, M., Bergdahl, N., Negrea, V., Oxley, E., Pham, P., Chong, S. W., & Siemens, G. (2024). A meta systematic review of artificial intelligence in higher education: A call for increased ethics, collaboration, and rigour. *International Journal of Educational Technology in Higher Education, 21*, Article 4. https://doi.org/10.1186/s41239-023-00436-z

Boud, D., & Falchikov, N. (2006). Aligning assessment with long-term learning. *Assessment & Evaluation in Higher Education, 31*(4), 399–413. https://doi.org/10.1080/02602930600679050

Chan, C. K. Y. (2023). A comprehensive AI policy education framework for university teaching and learning. *International Journal of Educational Technology in Higher Education, 20*, Article 38. https://doi.org/10.1186/s41239-023-00408-3

Chan, C. K. Y., & Hu, W. (2023). Students' voices on generative AI: Perceptions, benefits, and challenges in higher education. *International Journal of Educational Technology in Higher Education, 20*, Article 43. https://doi.org/10.1186/s41239-023-00411-8

Chi, M. T. H., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist, 49*(4), 219–243. https://doi.org/10.1080/00461520.2014.965823

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

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences, 111*(23), 8410–8415. https://doi.org/10.1073/pnas.1319030111

katyara, priya, Soomro, S., Naich, M., & Harjani, N. D. (2025). Artificial Intelligence in Education: Insights from Thematic Analysis on Personalization, Engagement, and Academic Performance. *Journal of Innovative Research*, *3*(3), 44-54. https://doi.org/10.54536/jir.v3i3.5849

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences, 103*, Article 102274. https://doi.org/10.1016/j.lindif.2023.102274

Kestin, G., Miller, K., Klales, A., Milbourne, T., & Ponti, G. (2025). AI tutoring outperforms in-class active learning: An RCT introducing a novel research-based design in an authentic educational setting. *Scientific Reports, 15*, Article 17458. https://doi.org/10.1038/s41598-025-97652-6

Kumar, P. (2024). Large language models (LLMs): Survey, technical frameworks, and future challenges. *Artificial Intelligence Review, 57*, Article 260. https://doi.org/10.1007/s10462-024-10888-y

Liang, W., Yuksekgonul, M., Mao, Y., Wu, E., & Zou, J. (2023). GPT detectors are biased against non-native English writers. *Patterns, 4*(7), Article 100779. https://doi.org/10.1016/j.patter.2023.100779

Lin, M. P.-C., Chang, D. H., Winne, P. H., Bowen, G. M., & Ryoo, J. (2026). Mapping generative AI in higher education teaching and learning: A scoping review. *International Journal of Educational Technology in Higher Education, 23*, Article 55. https://doi.org/10.1186/s41239-026-00630-9

Lo, C. K. (2023). What is the impact of ChatGPT on education? A rapid review of the literature. *Education Sciences, 13*(4), Article 410. https://doi.org/10.3390/educsci13040410

Long, D., & Magerko, B. (2020). What is AI literacy? Competencies and design considerations. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1–16). Association for Computing Machinery. https://doi.org/10.1145/3313831.3376727

Lyu, J.-Y. . (2026). The Investment Uncanny Valley: Narrative Realism, Cognitive Dissonance, and Behavioral Biases in Cryptocurrency Markets. *Journal of Innovative Research*, *4*(1), 10-17. https://doi.org/10.54536/jir.v4i1.5395

Marigocio, C. M. G. . (2026). Value and Cost of Using AI-powered Writing Assistants in Academic Writing: Basis for Crafting a Policy Brief. *Journal of Innovative Research*, *4*(1), 23-37. https://doi.org/10.54536/jir.v4i1.6746

Miao, F., & Cukurova, M. (2024). *AI competency framework for teachers*. UNESCO. https://www.unesco.org/en/articles/ai-competency-framework-teachers

Miao, F., & Holmes, W. (2023). *Guidance for generative AI in education and research*. UNESCO. https://doi.org/10.54675/EWZM9535

Miao, F., Shiohira, K., & Lao, N. (2024). *AI competency framework for students*. UNESCO. https://www.unesco.org/en/articles/ai-competency-framework-students

Minaee, S., Mikolov, T., Nikzad, N., Chenaghlu, M., Socher, R., Amatriain, X., & Gao, J. (2024). *Large language models: A survey* [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2402.06196

Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education, 31*(2), 199–218. https://doi.org/10.1080/03075070600572090

Perkins, M., Furze, L., Roe, J., & MacVaugh, J. (2024).

The Artificial Intelligence Assessment Scale (AIAS): A framework for ethical integration of generative AI in educational assessment. *Journal of University Teaching and Learning Practice, 21*(6). https://doi.org/10.53761/q3azde36

Salameh, W. (2026). What should students not yet delegate to AI? Toward a theory of cognitive delegation boundaries in higher education. *Frontiers in Education, 11*, Article 1982690. https://doi.org/10.3389/feduc.2026.1982690

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4

Tabassi, E. (2023). *Artificial intelligence risk management framework (AI RMF 1.0)* (NIST AI 100-1). National Institute of Standards and Technology. https://doi.org/10.6028/NIST.AI.100-1

Tertiary Education Quality and Standards Agency. (2023). *Assessment reform for the age of artificial intelligence*. https://www.teqsa.gov.au/guides-resources/resources/corporate-publications/assessment-reform-age-artificial-intelligence

UNESCO. (2022). *Recommendation on the ethics of artificial intelligence*. https://www.unesco.org/en/articles/recommendation-ethics-artificial-intelligence

University of Cape Town. (2025, July 15). *Introducing UCT's AI in Education Framework*. UCT News. https://www.news.uct.ac.za/article/-2025-07-15-introducing-ucts-ai-in-education-framework

University of Oxford. (2025). *AI use in summative assessment*. https://www.ox.ac.uk/about/how-we-are-run/policies-and-statements/policy-hub/ai-use-in-summative-assessment

Walters, W. H., & Wilder, E. I. (2023). Fabrication and errors in the bibliographic citations generated by ChatGPT. *Scientific Reports, 13*, Article 14045. https://doi.org/10.1038/s41598-023-41032-5

Weber-Wulff, D., Anohina-Naumeca, A., Bjelobaba, S., Foltýnek, T., Guerrero-Dib, J., Popoola, O., Šigut, P., & Waddington, L. (2023). Testing of detection tools for AI-generated text. *International Journal for Educational Integrity, 19*, Article 26. https://doi.org/10.1007/s40979-023-00146-z

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education—Where are the educators? *International Journal of Educational Technology in Higher Education, 16*, Article 39. https://doi.org/10.1186/s41239-019-0171-0

Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory Into Practice, 41*(2), 64–70. https://doi.org/10.1207/s15430421tip4102_2

Source and Time Scope	Source Type and Context	Design or Evidential Basis	Principal Findings or Contribution	Synthesis Use and Limitations
Bastani *et al.* (2025) 2020–2025 evidence	Empirical study High-school mathematics, Türkiye; nearly 1,000 students.	Classroom-randomised field experiment comparing unrestricted GPT assistance, a safeguarded tutor and a control.	Both AI conditions improved assisted practice. Unrestricted assistance reduced subsequent unassisted performance; tutor safeguards largely mitigated this harm.	Supports distinguishing assisted performance from independent competence and testing hints and safeguards. Secondary-school evidence requires cautious transfer to universities; safeguards did not establish superior unassisted learning.
Biggs (1996) Foundational exception	Conceptual pedagogical contribution Higher education.	Integration of constructivist learning and alignment of instructional objectives, activities and assessment.	Constructive alignment links intended learning outcomes with the activities that develop them and assessments that demonstrate achievement.	Supports aligning AI permissions and assessment with intended competence. Application to LLMs is an interpretation; the publication does not test AI-supported pedagogy.
Black and Wiliam (1998) Foundational exception	Research review Classroom formative assessment across educational settings.	Critical synthesis of research on assessment, feedback and learner involvement.	Evidence indicates that strengthening formative assessment can improve learning; feedback and student self-assessment are central concerns.	Supports formative feedback and opportunities to act on it. Benefits of AI-generated feedback cannot be inferred automatically from evidence on formative assessment generally.

Bond *et al.* **(2024) 2020–2025 evidence**	Meta systematic review AI in formal higher and continuing education.	Synthesis of 66 evidence reviews published from 2018 to July 2023.	Identifies adaptive systems and personalisation as prominent applications, alongside needs for stronger ethics, methodological rigour and contextual attention.	Supports critical appraisal and context-sensitive institutional evaluation. Reviews overlap and cover broader AI applications; they are not 66 independent experiments on LLMs.
Boud and Falchikov (2006) Foundational exception	Conceptual assessment contribution Higher education and learning beyond graduation.	Theoretical argument connecting assessment with participation in practice and future learning.	Assessment should develop students' capacity to judge their own work and sustain learning beyond immediate certification.	Supports student evaluation of generated outputs and long-term competence. It offers a rationale, rather than evidence that AI assistance produces these capabilities.
Chan (2023) 2020–2025 evidence	Empirical and framework-development study Hong Kong universities.	Quantitative and qualitative evidence from 457 students and 180 teachers and staff.	Proposes an AI education policy framework with pedagogical, governance and operational dimensions; highlights stakeholder participation.	Connects classroom design with privacy, accountability, infrastructure and training. Stakeholder evidence and framework development do not establish causal policy effectiveness or universal applicability.
Chan and Hu (2023) 2020–2025 evidence	Empirical perceptions study 399 undergraduate and postgraduate students in Hong Kong.	Cross-sectional survey examining familiarity, willingness, perceived benefits and concerns.	Students were generally positive about generative AI support but raised concerns about accuracy, privacy, ethics and personal development.	Supports student consultation and targeted AI literacy. Self-reported usefulness and intention are not measures of learning gains; context and sampling limit transferability.
Chi and Wylie (2014) Foundational exception	Theoretical framework with supporting evidence Laboratory and classroom learning.	Development of ICAP and examination of studies of note taking, concept mapping and self-explanation.	Differentiates passive, active, constructive and interactive engagement and predicts stronger learning with deeper cognitive engagement.	Supports explanation, critique and substantive interaction in AI-supported tasks. Chatbot turn counts do not establish interactive engagement; LLM-specific effects require testing.
Cotton *et al.* **(2024) 2020–2025 evidence**	Conceptual and practice-oriented discussion Higher education academic integrity.	Analysis of ChatGPT opportunities, misconduct risks and institutional responses.	Discusses threats to conventional assessment and strategies involving assessment design, guidance and responsible use.	Supports explicit assistance rules, integrity education and assessment reform. Recommendations are not causal evidence of effectiveness; 2024 is the issue year, with online publication in 2023.
Freeman *et al.* **(2014) Foundational exception**	Meta-analysis Undergraduate STEM education.	Analysis of 225 studies comparing active learning with traditional lecturing.	Active learning improved examination performance and reduced failure relative to conventional lecturing across the included STEM studies.	Provides background support for purposeful student participation. It predates contemporary LLMs and does not establish the effectiveness of AI activities or all non-STEM applications.

Kasneci *et al.* **(2023) 2020–2025 evidence**	Interdisciplinary position paper Educational applications of LLMs.	Conceptual examination of opportunities, limitations and educational responsibilities.	Identifies potential for personalised explanations and support alongside inaccuracies, bias, overreliance and the need for human oversight.	Supports verification, teacher judgment and pedagogical safeguards. It maps plausible applications and risks rather than estimating an intervention's causal learning effect.
Kestin *et al.* **(2025) 2020–2025 evidence**	Empirical learning-outcome study College physics, Harvard University, United States.	Randomised crossover trial comparing a purpose-built AI tutor with in-class active learning across two lessons.	The AI tutor produced higher immediate learning gains in the studied lessons; students also reported stronger engagement and motivation.	Supports carefully scaffolded tutoring in the tested setting. Short duration, subject, institution and tutor design limit claims about long-term retention, transfer or unrestricted chatbot use.
Liang *et al.* **(2023) 2020–2025 evidence**	Empirical detector evaluation Native and non-native English writing samples.	Comparison of detector classifications across writing samples and prompting conditions.	Tested detectors disproportionately misclassified non-native English writing; relatively simple changes could also evade detection.	Supports fairness safeguards and caution about detector-based misconduct judgments. Findings concern tested tools and samples, not every detector or future system.
Lo (2023) 2020–2025 evidence	Rapid evidence review Educational uses across subject domains.	Content analysis of 50 articles from the first three months after ChatGPT's release, December 2022–February 2023.	Reports possible instructional and tutoring uses, variable subject performance, inaccurate information and assessment challenges.	Supports early mapping of benefits and risks. The short observation window and early model generation limit applicability to later tools and durable learning outcomes.
Long and Magerko (2020) 2020–2025 evidence	Conceptual framework from literature synthesis AI literacy and human–computer interaction.	Interdisciplinary synthesis of competencies and learner-centred design considerations.	Defines AI literacy around effective interaction with and critical evaluation of AI, organising competencies in a conceptual framework.	Supports explicit student and staff capability development. The framework concerns AI broadly and is not an evaluation of an LLM curriculum or institutional training programme.
Miao and Cukurova (2024) 2020–2025 evidence	UNESCO competency framework Teachers; international reference.	Normative specification of teacher competencies and progression.	Sets out 15 competencies across five dimensions and three progression levels: Acquire, Deepen and Create.	Supports professional development in human-centred practice, ethics, AI foundations, pedagogy and professional learning. Five dimensions are not five levels; implementation effectiveness requires local evaluation.
Miao and Holmes (2023) 2020–2025 evidence	UNESCO policy guidance Education and research internationally.	Human-centred analysis and recommendations for regulation, ethical validation and pedagogical design.	Emphasises human agency, privacy, inclusion, institutional preparedness and educational suitability of generative AI.	Supports institutional rules, tool validation and human oversight. Normative guidance does not itself demonstrate learning gains or constitute a local legal compliance assessment.

Miao *et al.* **(2024) 2020–2025 evidence**	UNESCO competency framework Students; school-curriculum orientation.	Normative specification of learning objectives and developmental progression.	Identifies 12 competencies across four dimensions, with Understand, Apply and Create as three progression levels.	Offers concepts for responsible AI literacy and critical judgment. Transfer to higher education requires adaptation; it is not a university intervention study.
Nicol and Macfarlane-Dick (2006) Foundational exception	Conceptual model grounded in research Higher education feedback.	Reinterpretation of formative-assessment evidence through self-regulated learning.	Proposes seven principles of feedback that help students monitor performance and take control of learning.	Supports reflective use of AI feedback, self-assessment and revision. The model does not show that generated feedback is accurate or automatically develops self-regulation.
Sweller (1988) Foundational exception	Theory with modelling and experimental evidence Problem solving and schema acquisition.	Computational account and experiments examining processing demands during learning.	Conventional means–ends problem solving can consume capacity needed for schema acquisition; instructional demands affect lear ning.	Supports staged explanations and attention to prior knowledge. Applying cognitive load theory to LLM assistance is an extension requiring direct evaluation.
Tabassi (2023) 2020–2025 evidence	NIST risk management framework Cross-sector organisational use.	Voluntary framework for identifying, evaluating and managing AI risks.	Organises risk management around Govern, Map, Measure and Manage, with attention to trustworthy AI.	Supports institutional responsibilities, procurement, monitoring and incident response. It is not education-specific and does not measure teaching effectiveness or certify compliance.
TEQSA (2023) 2020–2025 evidence	Expert assessment guidance Australian higher education.	Expert collaboration and sector consultation on assessment reform.	Advocates assessment approaches that manage AI risks while preparing students to use AI appropriately.	Supports assessment redesign and credible demonstrations of learning. Guidance requires contextual implementation and evaluation; it is not a comparative trial of assessment strategies.
UNESCO (2022) 2020–2025 evidence	International ethical recommendation Cross-sector AI governance.	Normative principles and policy action areas; adopted in 2021, published in the cited 2022 edition.	Places human rights, dignity, transparency, fairness and human oversight at the centre of AI governance.	Supports equity, privacy and accountability dimensions of institutional policy. International ethical principles do not demonstrate implementation success or replace local institutional requirements.
University of Cape Town (2025) 2020–2025 evidence	Institutional framework announcement South African university.	Official account of framework development and consultation; announced 15 July 2025.	Describes three pillars: AI literacies, assessment integrity and innovation in teaching, learning and curriculum design.	Provides an African institutional example connecting pedagogy and governance. An announcement documents intentions and structures, not measured implementation or learning outcomes.

University of Oxford (2025) 2020–2025 evidence	Institutional assessment policy United Kingdom university.	Published requirements for assessment setters and students; public page updated 27 October 2025.	Requires explicit permissions, alignment of assessment with permitted assistance, equitable baseline provision and declarations of AI use.	Provides an example of bounded permissions and transparent assessment rules. A published policy is not evidence of effectiveness; transfer requires attention to local capacity and version dates.
Walters and Wilder (2023) 2020–2025 evidence	Empirical bibliographic audit ChatGPT-generated academic citations.	Verification of references generated by GPT-3.5 and GPT-4 against bibliographic records.	Documents fabricated references and errors in real references, despite improved performance in the newer tested model.	Supports checking source existence and bibliographic accuracy against originals. Results reflect tested models and prompts; they do not quantify every later system's error rate.
Weber-Wulff *et al.* (2023) 2020–2025 evidence	Empirical benchmark study AI-generated text detection in academic contexts.	Testing of 12 public tools and two commercial systems, including translated and obfuscated texts.	The tested tools were insufficiently accurate and reliable; obfuscation further weakened detection.	Supports triangulating misconduct evidence and avoiding decisive reliance on detector scores. Benchmark findings are version- and dataset-dependent, with rapidly changing tools.
Zawac-ki-Richter *et al.* (2019) Background exception	Systematic review AI in higher education before contemporary LLMs.	Synthesis of 146 articles covering 2007–2018.	Identifies profiling and prediction, assessment, adaptive systems and tutoring, with weak pedagogical grounding and limited critical ethical discussion.	Provides historical context and justification for educator-led integration. Outside the 2020–2025 window and not evidence about contemporary LLM effectiveness; retain only as an explicit background exception.
Zimmerman (2002) Foundational exception	Theoretical overview Self-regulated learning.	Explanation of cyclical learner processes involving forethought, performance and reflection.	Emphasises learners' active management of goals, strategies, monitoring and evaluation.	Supports independent attempts, learning logs and reflective verification of AI assistance. It supplies theory rather than evidence that AI interaction causes self-regulation.

APPENDIX 1: POPULATED EVIDENCE MATRIX Appendix 2: The Implementation Matrix translates Proposed Five-Stage Cycle for bounded AI assistance into practical guidance for the University Policy

Stage and learning purpose	Student activity and AI boundary	Evidence of learning	Institutional responsibility and quality assurance
1. Independent engagement Establish the student's initial understanding.	Complete an initial interpretation, solution, or argument independently. LLM assistance with the reasoning being assessed is outside the permitted scope at this stage.	An initial draft, calculation, concept map, or oral explanation that makes the student's reasoning visible.	**Course instructor** specifies the independent task in the assessment brief. **Programme team** checks that it aligns with the intended learning outcome and can be completed with the resources available to students.
2. Guided AI use — Obtain feedback or explore alternative explanations.	Use an approved LLM for a defined purpose, such as requesting feedback, testing a counterargument, or comparing explanations. The student retains responsibility for the assessed reasoning and discloses how AI was used.	A brief record of the assistance sought, the relevant output, and how it informed the student's next step.	**Course instructor** specifies permitted uses and disclosure requirements. **ICT and teaching support teams** arrange access or support an equivalent activity where direct interaction with an LLM is outside the intended outcome.

3. Verification — Evaluate the accuracy and relevance of AI outputs.	Check AI-generated claims, calculations, or references against course materials, primary evidence, or authoritative disciplinary sources. Explain which claims were accepted, revised, or rejected.	An annotated response, verification log, or corrected calculation that shows what was checked and the basis for each decision.	**Course instructor** assesses verification using explicit rubric criteria. **Library and teaching support teams** help students access suitable sources and develop verification skills.
4. Application — Demonstrate transfer of learning.	Apply the underlying concept to a new problem or context. Where independent competence is the intended outcome, complete the transfer task without LLM assistance.	An independently completed analysis, problem solution, practical demonstration, or oral defence scored against the intended outcome.	**Programme team** moderates the transfer task and rubric. **Quality assurance unit** reviews whether the assessment yields comparable evidence across delivery and access routes.
5. Reflection — Evaluate the contribution and limits of AI assistance.	Explain how the initial reasoning changed, which suggestions influenced the revision, and what the student can now do independently. Complete the reflection in the student's own words.	A short reflective statement linked to the initial attempt, verification record, and application task.	**Course instructor** reviews reflections for recurring misconceptions and access barriers. **Programme team** uses those findings to revise task instructions, rubrics, and student support.