SAMYRAD 2026

SMART ALGORITHMS & MODELS: YOUR REACH TO ADVANCED DIALOGUES · 5–6 OCTOBER 2026 · SEVILLE, SPAIN

SAMYRAD OPEN-ACCESS PROCEEDINGS | PUBLISHER: INTELLIGENT NUDGE

PAPER • OPEN ACCESS

From Adoption to Human Judgement: Responsible Generative AI Integration in Higher Education

Kurt Naicker[1*], Katlego Thipe[2]

*Corresponding author: kurtn@uj.ac.za

[1]Department of Commercial Accounting, University of Johannesburg, Johannesburg, South Africa
[2]Department of Accountancy, University of Johannesburg, Johannesburg, South Africa

Article DOI: https://doi.org/10.68286/samyrad2026.292

ABSTRACT

Generative artificial intelligence (GenAI) is transforming higher education faster than institutional pedagogy and governance can adapt. Although recent scholarship documents opportunities and risks, the relationships among pedagogical use, personalisation, governance and human cognitive capability remain fragmented. This paper presents a structured qualitative evidence synthesis using directed content analysis of 28 publications from 2024 to 2026, comprising systematic reviews, narrative or conceptual reviews, and selected empirical studies. Four predefined domains guided the analysis: AI-enhanced teaching and learning, personalised and adaptive learning, ethical governance, and AI literacy and critical thinking. The selected literature consistently reports potential benefits for content generation, formative feedback and personalisation, while also identifying fragmented governance, persistent equity constraints and limited empirical testing of safeguards against cognitive overreliance. The paper proposes a human-centred conceptual framework in which pedagogical applications are enabled by AI literacy, bounded by ethical governance and conditioned by institutional context. Critical thinking, professional judgement and graduate readiness are positioned as central human outcomes. This is a conceptual and theoretical contribution rather than an empirical evaluation of GenAI's effects, and the proposed framework has not yet been empirically tested. The framework is conceptual rather than causal and provides propositions for future empirical testing and practical guidance for universities, lecturers, students, policymakers and curriculum developers.

KEYWORDS Generative artificial intelligence, higher education, directed content analysis, responsible AI, AI literacy, critical thinking, educational governance.

How to cite: Naicker, K., & Thipe, K. (2026). From adoption to human judgement: Responsible generative AI integration in higher education. SAMYRAD 2026 Open-Access Proceedings. Intelligent Nudge. https://doi.org/10.68286/samyrad2026.292

Introduction

AI Is Transforming Higher Education

Generative artificial intelligence (GenAI) has moved from a peripheral technology to a central feature of the higher-education landscape within a remarkably short period. The public release of large language models such as ChatGPT precipitated what several of the studies reviewed here describe as a rapid, largely unplanned wave of institutional and individual adoption [1], [2]. Universities now encounter GenAI simultaneously as a teaching aid, an assessment risk, a research tool, and an object of policy concern — often before any of these roles has been deliberately defined. This convergence of opportunity and disruption forms the starting point for the present review: GenAI is not a single, bounded innovation to be evaluated once, but a diffuse capability whose implications for pedagogy, equity, ethics and cognition are still being worked out across a fast-growing and fragmented body of literature.

The opportunities are considerable. Reviewed studies report that GenAI can generate differentiated instructional content, provide immediate formative feedback, model learner needs dynamically, and free educator time from routine tasks [1], [3], [4]. Yet the same literature documents a parallel set of concerns around privacy, bias, academic integrity, governance capacity, and the risk that convenience gradually substitutes for independent thought [5], [6], [7]. The result is a literature that is expanding faster than it is being synthesised.

Existing Literature

Recent reviews have begun to map this terrain, typically clustering around one of three concerns. A first cluster addresses teaching and assessment, documenting how GenAI reshapes content creation, feedback and the educator's role in the classroom [8], [9]. A second cluster addresses ethics and governance, cataloguing the privacy, bias, integrity and regulatory risks associated with institutional adoption [10], [11], [12]. A third, smaller cluster addresses AI literacy, examining whether students and staff are equipped to use these tools critically rather than uncritically [13], [14]. Each cluster has produced valuable, largely self-contained syntheses. What has received comparatively little attention is how these three conversations — pedagogy, governance, and cognition — relate to one another as GenAI is integrated into everyday academic practice.

The Gap

Existing reviews have catalogued opportunities and risks associated with AI adoption in higher education. However, comparatively less attention has been devoted to synthesising how these issues collectively reshape the relationship between pedagogy, governance, and human cognitive capability. Studies of AI-enhanced teaching rarely engage with the governance literature that constrains its safe deployment; studies of ethical governance rarely engage with the pedagogical evidence on whether generated content actually reaches higher-order learning outcomes; and studies of AI literacy are frequently proposed as a standalone remedy without being connected back to either the pedagogical or governance conversations that

create the need for it in the first place. This fragmentation matters because responsible AI integration is, by definition, a cross-cutting institutional problem rather than a series of independent technical or ethical questions.

Research Objective

This paper conducts a directed qualitative content analysis of recent literature to identify the dominant themes shaping responsible GenAI integration in higher education and to develop an integrated conceptual account of their relationships. Rather than treating pedagogy, personalisation, governance and literacy as independent concerns, the paper examines how these domains interact in shaping critical thinking, professional judgement and graduate readiness. The study addresses the following research questions:

RQ1: What recurring pedagogical, personalisation, governance and literacy-related themes characterise contemporary literature on responsible GenAI integration in higher education?

RQ2: How do these themes interact in shaping critical thinking, professional judgement and graduate readiness within AI-enabled learning environments?

Literature review

This section situates the analysis conceptually. It does not analyse the 28 reviewed studies individually — that task is undertaken in Section IV — but establishes the definitions and conceptual boundaries the analysis subsequently applies.

Artificial Intelligence in Higher Education

Artificial intelligence in higher education encompasses a broad family of technologies, from earlier rule-based intelligent tutoring systems and predictive learning analytics through to contemporary large language models [1], [15]. What distinguishes the current period from earlier waves of educational technology is scale and accessibility: AI capability that was once confined to specialist systems is now available directly to students and staff through consumer-facing interfaces, collapsing the distance between institutional adoption decisions and individual, often unsanctioned, use [2].

Generative AI

Generative AI (GenAI) refers specifically to models capable of producing novel text, code, images or other content in response to a prompt, as distinct from earlier, predominantly classificatory or predictive applications of AI in education [3]. In the higher-education context, this generative capability is what enables GenAI to function simultaneously as a content-creation tool for educators, a study aid for students, and an assessment risk for institutions — three roles that earlier, narrower AI applications rarely combined [9], [16].

Responsible AI

Responsible AI, as used across the reviewed governance literature, refers to the deliberate design of institutional structures policy, oversight, training and technical safeguards intended to ensure that AI benefits are realised without disproportionate harm to privacy, fairness, or academic integrity [12], [17]. The concept is normative rather than purely technical: it presumes that responsibility is distributed across developers, institutions, educators and students, and that governance mechanisms such as human-in-the-loop review or risk-tiering are only meaningful when embedded in institutional practice rather than left as abstract principle [11].

AI Literacy

AI literacy denotes the capacity of learners and educators to understand, evaluate and use AI tools critically — including recognising their limitations, biases and appropriate boundaries of use — rather than treating AI output as authoritative by default [13], [10]. Across the literature, AI literacy is positioned simultaneously as a curricular outcome in its own right and as an enabling condition for the safe use of GenAI elsewhere in the curriculum, which is part of why it recurs across multiple thematic clusters in the analysis that follows.

Critical Thinking

Critical thinking, in this literature, is generally treated as the higher-order cognitive capacity that GenAI use may either scaffold or erode, depending on how it is deployed [18], [19]. Where GenAI is used to support reflection, iteration and evaluation of ideas, several studies report a positive relationship with metacognitive development; where it is used as a substitute for independent reasoning, the same literature reports a risk of cognitive offloading and de-skilling [6], [7]. This unresolved tension is a central concern of the present review and is returned to directly in the Results and Discussion.

Human Agency, Professional Judgement and Graduate Readiness

Human agency refers to the capacity of educators and learners to retain meaningful control over the purposes, processes and consequences of AI-supported activity. In this paper, professional judgement is treated as the disciplined application of knowledge, critical evaluation and contextual reasoning for which human actors remain accountable. Graduate readiness extends beyond operational familiarity with AI tools to include the capacity to question outputs, recognise uncertainty, justify decisions and act responsibly within disciplinary and professional settings. These concepts provide the human-centred outcomes against which the responsible integration of GenAI is assessed.

Methodology

Research Philosophy

The study adopted an interpretivist orientation because its purpose was to synthesise how contemporary publications construct and connect the pedagogical, governance and cognitive dimensions of GenAI use. A structured qualitative evidence-synthesis design was used, with directed content analysis providing the analytical method.

Review Design

The review is a structured qualitative synthesis of publication-level evidence rather than a systematic review of primary studies. The corpus comprises systematic reviews, narrative or conceptual reviews and selected empirical studies. The publication, rather than the participant sample within an underlying primary study, was the unit of analysis. This design was selected to integrate themes across an already substantial review-level literature and to identify relationships that are often examined separately. The review does not claim exhaustive retrieval. Several included publications synthesise overlapping bodies of primary research; consequently, recurrence across publications indicates thematic consistency in the selected corpus rather than the number of independent empirical datasets supporting a claim.

Literature Search Strategy and Study Selection

A structured, iterative search was conducted through the University of Johannesburg Library environment using Scopus, Web of Science, SpringerLink, ScienceDirect and Google Scholar. Search activity focused on English-language scholarship published from 2022 to 2026 and used combinations of terms relating to artificial intelligence, generative artificial intelligence, higher education, teaching and learning, personalisation, governance, AI literacy and critical thinking. The principal search expression retained in the review audit trail was:

("Artificial Intelligence" OR "AI" OR "Generative Artificial Intelligence" OR "GenAI" OR "ChatGPT") AND ("Higher Education" OR universit OR college*) AND (teaching OR learning OR assessment OR pedagogy OR education OR "AI literacy" OR "critical thinking")*

Additional searches were undertaken around the four analytical domains: AI-enhanced teaching and learning; personalised and adaptive learning; ethical and governance considerations; and AI literacy and critical thinking. Because database-specific search dates, field restrictions, result counts and all supplementary search strings were not recorded contemporaneously, the process cannot be represented as a fully reproducible systematic search. The paper therefore reports the review transparently as a structured and purposive evidence synthesis rather than retrospectively reconstructing unrecorded methodological details.

Publications were retained when they addressed AI or GenAI in higher education and contributed substantively to at least one analytical domain. Peer-reviewed empirical studies, systematic reviews, narrative reviews and conceptual studies were eligible.

Publications were excluded when they focused exclusively on primary or secondary education, addressed technical model development without educational application, were non-scholarly commentary, were not available in English or fell outside the conceptual scope.

The working literature matrix contained 28 publications retained for synthesis, published between 2024 and 2026. Selection and eligibility assessment were completed by the first author using the stated criteria and a structured matrix. A PRISMA flow diagram is not presented because complete database-level identification, deduplication and exclusion-reason records were not maintained. This limitation is acknowledged to avoid presenting retrospective estimates as observed review data.

Data Extraction

The 28 publications were extracted into a structured literature matrix capturing authorship, year, geographical scope, journal, research aim, design, sample, AI technology, primary theme, findings, limitations, practical implications, future-research recommendations, quality-control observations and a provisional evidence-weight classification.

The evidence-weight classification was used as a sensitivity aid rather than a formal risk-of-bias score. Publications were considered higher confidence when they reported transparent retrieval and eligibility procedures, independent screening or coding, and study-level appraisal. Publications with incomplete corpus reporting, internal inconsistencies, miscitation or weak source integrity were treated as lower-confidence or contextual evidence. The synthesis prioritised patterns supported by stronger sources and used lower-confidence publications only for corroboration or context.

Directed Qualitative Content Analysis

Directed qualitative content analysis was applied in accordance with Hsieh and Shannon [20]. The four predefined domains were derived from the research objective and preliminary scoping of the literature: (1) AI-enhanced teaching and learning, (2) personalised and adaptive learning, (3) ethical and governance considerations, and (4) AI literacy and critical thinking.

Analysis proceeded in three stages. First, relevant meaning units were deductively assigned to one or more predefined domains. Second, descriptive codes were generated from the research aims, findings, limitations and recommendations recorded in the matrix. Third, semantically related codes were grouped inductively into subcategories and higher-order interpretive sub-themes within each domain. A publication was counted as contributing to a category when its extracted findings or recommendations contained a substantive meaning unit relevant to that category.

Frequency (f) indicates the number of publications contributing to a category, not the number of code occurrences, participants or independent primary datasets. Frequencies are descriptive indicators of recurrence within the selected corpus and should not be interpreted as prevalence estimates, effect sizes or evidence-quality scores.

Trustworthiness

Trustworthiness was supported through a structured extraction matrix, explicit category definitions and an audit trail linking each reported category to contributing publications. Coding was conducted by the first author and reviewed conceptually with the second author during manuscript development; however, independent double-coding and a formal inter-coder reliability statistic were not completed.

Interpretive findings should therefore be understood as analytically transparent but not independently reliability-tested.

The review has four additional limitations. First, the search was structured and purposive rather than fully systematic, and complete database-level search logs were not retained. Second, the corpus combines heterogeneous publication types and methodological quality. Third, several included reviews may rely on overlapping primary studies, which can amplify apparent convergence. Fourth, English-language and review-level evidence may underrepresent local, non-English and emerging Global South scholarship. These limitations constrain claims of exhaustiveness and causality but do not prevent the review from identifying recurring conceptual relationships within the selected corpus.

These four limitations are interrelated and should be read together when interpreting the Results: because selection was completed by a single researcher rather than through independent double screening, because overlapping primary-study corpora across included reviews can inflate the apparent number of studies “agreeing” on a pattern without adding independent evidence, and because the corpus mixes review types of differing rigour, the frequency counts reported in Section IV should be read as indicators of thematic consistency within this particular, purposively curated corpus — not as independent replication counts, effect-size evidence, or a claim generalisable beyond the reviewed literature.

Results

The following sub-sections report the axial coding summary and narrative pattern (agreements, contradictions and gaps) for each of the four predefined themes.

Theme 1: AI-Enhanced Teaching and Learning

Fourteen of the 28 records address how GenAI reshapes instructional delivery, feedback and the educator's role. Open coding of aims and findings yielded six axial categories (Table 1), synthesised into three sub-themes: GenAI as a content-and-feedback engine; the evolving educator role; and framework-guided versus ad hoc implementation. There is strong agreement that GenAI functions as a generative co-producer of instructional content and a first layer of formative feedback: curricula, quizzes, simulations and differentiated materials [1], [3], automated marking and feedback that supports self-regulated learning [4], and reduced routine

workload freeing educator time for higher-order tasks [9], [21]. This convergence extends across review types — conceptual, PRISMA-systematic, and mixed-method — suggesting the finding is robust to methodological variation. A contradiction emerges around pedagogical depth. Several studies report measurable gains in engagement, motivation and performance attributable to GenAI-enhanced teaching [2], [4], yet Mere [3] explicitly flags weak pedagogical depth in generated content, and Rodríguez-Ortiz et al. [15] find that GenAI-supported learning analytics show weak pedagogical grounding and limited educator involvement. Belkina et al. [8] is the clearest counter-example, demonstrating through Laurillard's Conversational Framework, SAMR and TPACK that GenAI use can progress from mere substitution to genuine pedagogical transformation — but only when implementation is explicitly framework-guided. The tension, therefore, is not whether GenAI can enhance teaching, but whether most reported enhancement is technology-led rather than pedagogy-led. The clearest gap is the scarcity of studies anchoring GenAI integration in an explicit instructional-design framework: only one of fourteen contributing studies [8] does so directly. Domain-specific applications beyond writing and art, and classroom-embedded evaluations of teacher-AI collaboration, remain rare relative to the volume of review-level synthesis.

Theme 2: Personalized and Adaptive Learning

Ten records centre on adaptive and personalised learning mechanisms, their empirical effectiveness, and the equity conditions under which benefits are realised. Six axial categories (Table 2) synthesise into three sub-themes: mechanisms of personalisation; evidence of effectiveness, unevenly theorised; and persistent equity, quality and implementation gaps.

Table 1. Axial coding summary — AI enhanced teaching and learning

Axial Category	Illustrative Open Codes	Contributing Studies	f
Automated content, curriculum & assessment generation	curriculum drafting; quiz/simulation generation; differentiated materials; process-based assessment design	[1], [3], [9], [18], [21], [22]	**6**
Automated & blended feedback mechanisms	AI-generated feedback; instructor–AI feedback integration; formative feedback loops; feedback for metacognition	[1], [4], [9], [15], [16], [21]	**6**
Educator role transformation & workload reduction	facilitation over transmission; reduced routine workload; human-AI collaboration; teacher optimisation	[1], [2], [4], [9], [13], [17], [18], [21]	**8**
Learning analytics for prediction & engagement	dropout prediction; engagement/sentiment analytics; scaffolding via analytics	[15]	**1**
Pedagogical-framework-guided implementation	SAMR substitution-to-transformation; TPACK alignment; Laurillard's Conversational Framework	[8]	**1**
Domain-specific applications (art; academic writing)	creative scaffolding; cultural-heritage content; ideation and outlining support; grammar/revision assistance	[16], [22]	**2**

Note: f denotes the number of contributing publications, not prevalence or evidence weight, Method section, Directed Qualitative Content Analysis.

Table 2. Axial coding summary — personalized and adaptive learning

Axial Category	Illustrative Open Codes	Contributing Studies	f
Adaptive pathways, pacing & difficulty adjustment	dynamic sequencing; difficulty calibration; individualised learning pathways; early-warning triggers	[1], [3], [23], [24], [25]	**5**
Learner modelling & profiling	dynamic learner profiles; multimodal learner modelling; emotion recognition; algorithmic personalisation engines	[3], [15], [24], [25]	**4**
Empirical effectiveness evidence	performance gains (15–25%); engagement gains (30–40%); retention and accessibility improvements	[23], [24], [26]	**3**
Cognitive-level & theoretical alignment	Bloom's taxonomy mapping; constructivist grounding; PBL/STEAM/UDL alignment	[19], [26]	**2**
Quality, multimodality & workload of generated resources	expert-review evaluation; mixed-methods quality checks; educator workload of resource curation	[15], [27]	**2**
Equity, access & implementation barriers	infrastructure cost; digital-divide access gaps; institutional readiness; cultural adaptation	[3], [23], [24], [26], [28]	**5**

Note: f denotes the number of contributing publications, not prevalence or evidence weight, Method section, Directed Qualitative Content Analysis.

The selected publications show their clearest agreement on mechanism and reported outcome: adaptive pathway sequencing, pacing and learner profiling are consistently

described as the operational core of AI personalisation [1], [3], [25], and where effectiveness is quantified, the direction of effect is uniformly positive — one included review reports ranges of 15–25% for performance and 30–40% for engagement across its selected studies [23]; these ranges are not pooled effect estimates. Fortuna et al. [24] similarly reports favourable patterns for retention and accessibility.

A contradiction surfaces between the maturity claimed for multimodal personalisation and the maturity observed in practice. Chakraborty [1] and Mere [3] present multimodal learner modelling (including emotion recognition) as an active capability, whereas Yuan et al. [27], examining GenAI-generated personalised resources directly, find that dynamic multimodal resources are comparatively rare and that immediate textual feedback and language support dominate actual implementations. A second tension concerns theoretical grounding: Peng and Li [26] find that constructivist theory is frequently implicit rather than explicit, and that when learning theory is applied deliberately it measurably improves pedagogical alignment. Equity is the most consistently under-operationalised construct in this theme: Peng and Li [26] note it is poorly operationalised even where nominally addressed, and geographic concentration is pronounced — 82.8% of that review's evidence base originates from Asia. Combined with barriers reported by Mere [3], Yuensook et al. [23] and Coman et al. [28], this indicates a gap in equity-focused, Global-South-inclusive evaluation of adaptive learning systems.

Theme 3: Ethical and Governance Considerations

Fourteen records — the largest cluster in the matrix — address ethical risk and governance, reflecting the field's current preoccupation with responsible deployment. Seven axial categories (Table 3) synthesise into three sub-themes: core, recurrent ethical risk domains; governance fragmentation and the institutional response gap; and emerging and underexamined costs.

The selected publications show substantial thematic agreement that privacy, algorithmic bias and academic integrity constitute the dominant, recurring ethical risk triad, appearing independently across conceptual reviews [1], systematic reviews [5], [11], and regionally focused studies [10]. This recurrence across geographically and methodologically diverse publications indicates a stable pattern in the selected corpus: the risk domains themselves are not contested, even though their proposed remedies differ.

Table 3. Axial coding summary — ethical and governance considerations

Axial Category	Illustrative Open Codes	Contributing Studies	f
Privacy & data governance	data protection; surveillance risk; consent; data minimisation	[1], [3], [5], [6], [11], [12], [14], [17]	**8**
Bias, fairness & algorithmic equity	algorithmic bias; cultural/language bias; unequal access to benefits; fairness auditing	[1], [3], [5], [12], [14], [17], [18], [28]	**8**
Academic integrity, authorship & misconduct	plagiarism; ghostwriting; authorship ambiguity; contract-cheating risk	[5], [6], [7], [9], [10], [11], [14], [16], [18]	**9**
Transparency, accountability & human oversight	explainability (XAI); human-in-the-loop review; risk-tiering; accountability structures	[1], [12], [14], [17]	**4**
Regulatory & institutional policy fragmentation	fragmented/reactive policy; absent institutional frameworks; participatory governance gaps	[5], [7], [10], [11], [13], [28]	**6**
Contextual & regional readiness constraints	infrastructure gaps; Global South readiness; policy-capacity constraints	[10]	**1**
Environmental & sustainability costs	compute/energy footprint; underexamined environmental cost of GenAI use	[7]	**1**

Note: f denotes the number of contributing publications, not prevalence or evidence weight, Method section, Directed Qualitative Content Analysis.

A contradiction exists between the sophistication of proposed governance mechanisms and their reported operational reality. Alfiras et al. [12] articulate concrete mechanisms — risk-tiering, human-in-the-loop review, audit trails and authentic assessment redesign — yet the same study concedes that existing institutional governance is often abstract and lacks operational detail, a tension echoed by Pireci Sejdiu et al. [7], who find that adoption outpaces governance, and by Kaşarcı et al. [11], who describe governance as reactive rather than proactive. A second, contextual contradiction is that readiness constraints reported by Owidi et al. [10] for Kenyan and wider African institutions are largely invisible in the Global-North-centred governance literature.

The most striking gap is environmental cost: of fourteen contributing studies, only Pireci Sejdiu et al. [7] raises GenAI's computational footprint as a governance concern, despite privacy and bias each drawing attention from eight studies. Empirical evaluation of governance-mechanism effectiveness is likewise absent from the corpus.

Theme 4: AI Literacy and Critical Thinking

Eleven records address the cognitive and literacy dimensions of GenAI use — the capacity of educators and learners to engage with these tools critically and productively. Four axial

categories (Table 4) synthesise into three sub-themes: dual potential (cultivation versus erosion of thinking); literacy as the primary mitigation mechanism; and underdeveloped higher-order engagement.

Table 4. Axial coding summary — AI literacy and critical thinking

Axial Category	Illustrative Open Codes	Contributing Studies	f
Critical thinking & metacognition cultivation	metacognitive scaffolding; reflective feedback loops; process-based assessment supporting reasoning	[1], [8], [13], [18]	**4**
Overreliance, cognitive offloading & de-skilling risk	automation bias; reduced critical reflection; dependence on AI output; erosion of independent reasoning	[6], [7], [9], [18]	**4**
Literacy development for students & educators	AI-literacy curricula; ethical-literacy training; educator upskilling; transparent co-writing practices	[3], [6], [10], [11], [12], [13], [14]	**7**
Higher-order cognitive skill gaps	limited evidence for analysis/evaluation/creation; concentration on lower-order recall and comprehension	[19]	**1**

Note: f denotes the number of contributing publications, not prevalence or evidence weight, Method section, Directed Qualitative Content Analysis.

Agreement centres on the dual-edged nature of GenAI's cognitive impact. Several studies find that, when purposefully scaffolded, GenAI use can support metacognition and reflective reasoning [1], [18], [13], while a comparably sized set of studies documents overreliance and reduced critical reflection as a realised risk rather than a hypothetical one [6], [7], [9].A contradiction worth surfacing is that the overreliance/de-skilling literature largely assumes GenAI tools are cognitively capable enough to be worth offloading to, whereas Latif et al. [19] — the only study in the corpus to map AI personalisation explicitly against Bloom's taxonomy — finds that most tools concentrate on remembering, understanding and applying, with limited evidence of engagement at the analysis, evaluation or creation levels. The offloading risk documented elsewhere may therefore apply more to foundational tasks than to complex reasoning, a distinction the reviewed literature does not yet make explicit.

Literacy training is the most frequently recommended mitigation across the matrix, cited by seven of eleven contributing studies — yet none report an empirical evaluation of its effect on critical-thinking outcomes. This is the theme's central gap, and higher-order cognitive skill development remains markedly under-researched.

Discussion

The Results reported in the Result Section are organised by coding theme; the Discussion instead organises the same evidence around three overarching arguments, moving from what the literature shows toward what it means for responsible AI integration in higher education.

Technology Is Advancing Faster Than Pedagogy

Themes 1 and 2 together (Results Section, Themes 1 and 2) point to a single underlying pattern: the field evaluates GenAI primarily on technological and behavioural outcomes — did engagement rise, did output improve — rather than on whether its use is pedagogically designed. Only Belkina et al. [8] demonstrates, through an explicit instructional-design framework, that GenAI use can progress to genuine transformation rather than substitution Adoption is therefore running ahead of instructional-design maturity, and the multimodal and equity claims noted in the Result Section, Theme 2: Personalized and Adaptive Learning, remain more aspirational than realised.

Governance Remains Reactive

Theme 3's central finding (Result Section, Theme 3: Ethical and Governance Considerations) is not disagreement about risk but a persistent gap between proposed governance mechanisms [12] and their operational reality, described elsewhere in the corpus as abstract, fragmented and reactive [7], [11]. This is consistent with policy formation generally lagging diffusion of a general-purpose technology, particularly where individual adoption can outpace institutional decision-making. Readiness constraints in under-resourced systems [10] being largely absent from the dominant governance literature further indicates that reactive governance is compounded, not uniform, where institutional capacity is already limited.

The Future Challenge Is Preserving Human Judgement

If pedagogy is not yet keeping pace with technology (V-A) and governance is not yet keeping pace with risk (V-B), the residual safeguard in both cases is the individual's own critical judgement — precisely the capacity Results Section, Theme 4, reports as most uncertain. Literacy training is proposed almost universally as the remedy, yet no study empirically evaluates whether it protects critical-thinking outcomes, and current tools rarely engage higher-order cognition in the first place [19]. The field has arrived at a proposed solution without establishing that it addresses the actual mechanism of risk; preserving human judgement, not merely expanding AI capability or policy, is the more durable challenge.

Theoretical contribution

Existing literature frequently treats AI-enhanced teaching, personalised learning, ethical governance and AI literacy as separate areas of inquiry. The present synthesis instead proposes that responsible GenAI integration in higher education should be understood as an interacting, human-centred system. This represents the paper's principal theoretical contribution.

The novelty of this contribution lies not in the four domains themselves, each of which is already documented separately in the reviewed literature, but in the explicit, directional specification of how they relate. Prior syntheses typically treat pedagogy, personalisation, governance and literacy as parallel concerns; this framework instead specifies that AI literacy and critical thinking function as a necessary enabling condition between pedagogical application and human-centred outcome (Propositions 1 and 3), that governance operates as a moderator rather than a downstream compliance step (Proposition 2), and that graduate readiness is contingent on retained human responsibility rather than an automatic by-product of adoption (Proposition 4). It is this set of testable, directional relationships not the underlying themes in isolation — that constitutes the paper's original contribution.

Within the proposed framework, AI-enhanced teaching and personalised learning constitute the main pedagogical applications of GenAI. AI literacy and critical thinking function as enabling human capabilities that influence whether students and educators engage with GenAI reflectively or become excessively dependent on its outputs. Ethical governance, including privacy, fairness, transparency, accountability and meaningful human oversight, surrounds and moderates the use of GenAI across the educational system. The effectiveness and equity of these relationships are further conditioned by infrastructure, institutional capacity, disciplinary requirements and sociocultural context.

The intended outcomes of responsible integration are not technology adoption or efficiency alone. The framework positions professional judgement and graduate readiness as central human outcomes. These outcomes are not automatic consequences of GenAI use; they depend on whether institutions design learning environments in which people remain responsible for evaluating information, recognising uncertainty, explaining decisions and exercising context-sensitive judgement.

The framework is conceptual and does not represent a validated causal sequence. Instead, the synthesis generates five propositions for future empirical investigation:

Proposition 1 (Pedagogical alignment): GenAI is more likely to enhance learning when its use is aligned with explicit learning outcomes, disciplinary requirements and appropriate instructional design.

Proposition 2 (Governance moderation): Ethical governance moderates the relationship between GenAI adoption and educational benefit by constraining unacceptable risks, clarifying accountability and preserving meaningful human oversight.

Proposition 3 (AI literacy and cognitive engagement): AI literacy influences whether engagement with GenAI strengthens reflective learning and critical thinking or encourages cognitive offloading and excessive dependence on generated outputs.

Proposition 4 (Human responsibility and graduate readiness): Graduate readiness depends on retaining human responsibility for evaluating evidence, recognising uncertainty, justifying conclusions and making context-sensitive decisions.

Proposition 5 (Contextual conditions): Infrastructure, institutional capacity, disciplinary context and sociocultural conditions shape the effectiveness, accessibility and equity of the relationships proposed in the framework.

Ethical governance surrounds and moderates the pedagogical use of GenAI, recognising that principles of privacy, fairness, transparency, accountability and meaningful human oversight permeate all stages of AI adoption rather than emerging only after implementation. AI literacy and critical thinking are positioned as enabling human capabilities that support reflective engagement with GenAI and reduce the risk of cognitive offloading and excessive dependence on AI-generated outputs. Professional judgement and graduate readiness represent the intended human-centred outcomes, while infrastructure, institutional capacity, disciplinary context and sociocultural conditions determine the feasibility, effectiveness and equity of responsible GenAI integration across higher education settings. Critically, this influence runs in both directions: observed outcomes for professional judgement and graduate readiness are expected to feed back into and recalibrate ethical-governance mechanisms over time, consistent with the framework's characterisation as an interacting rather than linear system (Fig. 1).

Practical implications

Universities

Institutions should treat GenAI governance as an operational, not merely declarative, function — resourcing risk-tiering, human-in-the-loop review and audit mechanisms rather than issuing policy statements alone [11], [12].

Institutions in resource-constrained contexts should be supported through sector-level or regional coordination, given that readiness constraints are compounded, not merely repeated, where infrastructure and policy capacity are already limited [10].

Lecturers

Educators are best placed to close the pedagogy gap identified by adopting explicit instructional-design rather than relying on ad hoc or tool-led adoption [8]. Assessment design should assume GenAI availability and shift toward process-based and authentic assessment that rewards demonstrated reasoning over polished output [12].

Students

Students require explicit AI-literacy instruction that goes beyond tool operation to include critical evaluation of GenAI output, recognition of hallucination and bias, and awareness of when independent reasoning — rather than AI assistance — is the more appropriate response to a task [13], [14].

Policy Makers

National and sector regulators should prioritise closing the gap between proposed governance mechanisms and their operational detail identified in the Discussion Section of this article, and should extend governance guidance to explicitly address environmental and sustainability costs, currently

addressed by only one study in this corpus [7].

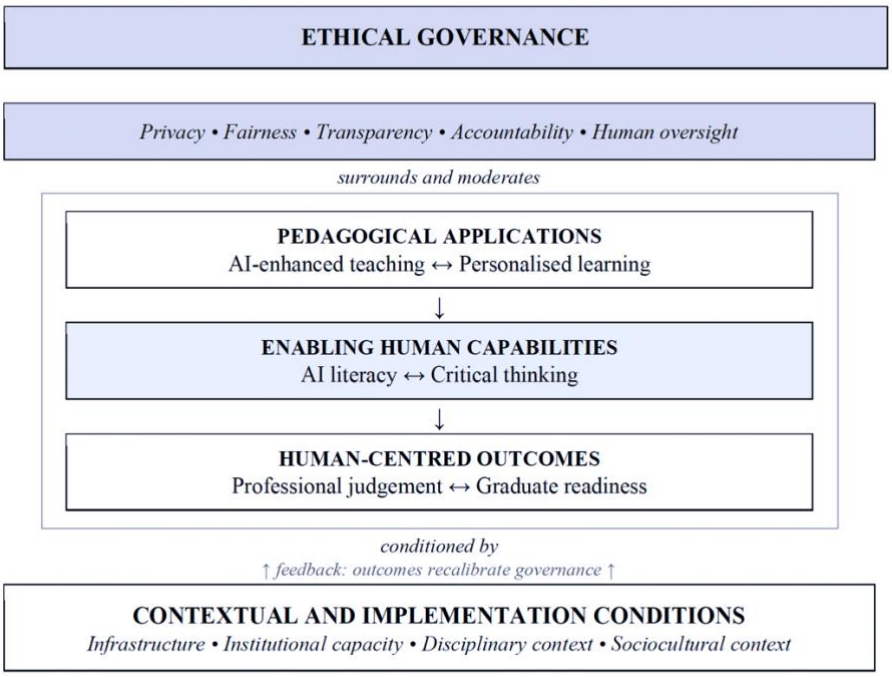


Fig. 1. visually represents the proposed human-centred framework and illustrates how the five propositions collectively conceptualise responsible GenAI integration as an interacting system rather than a linear process.

Curriculum Developers

Curriculum design should treat AI literacy and critical-thinking development as integrated rather than separate outcomes, embedding literacy instruction within disciplinary content rather than as a standalone module, and should build in evaluation mechanisms so that, unlike the studies reviewed here, literacy interventions can be assessed for their actual effect on critical-thinking outcomes rather than assumed to be effective.

Operationalising context: future work should treat "sociocultural context" in Proposition 5 as a measurable rather than background variable — for example, through indicators such as institutional GenAI-access rates, the language of instruction relative to the language(s) supported by deployed tools, and locally negotiated norms around acceptable AI assistance in assessment. Embedding such indicators into future empirical tests of the framework would allow context to function as an explanatory variable rather than a residual category.

Conclusion

This review synthesised 28 contemporary publications into an integrated conceptual account of responsible GenAI integration in higher education, addressing RQ1 (recurring themes) and RQ2 (their interaction). This matters because higher education's central claim to value the development of graduates capable of independent, disciplined professional judgement is precisely the capacity most exposed by the gaps this review identifies. Treating AI-enhanced teaching, personalised learning, ethical governance and AI literacy as four independent concerns, as much of the existing literature currently does, risks addressing each symptom without addressing the underlying system that connects them.

Future research should prioritise three directions: empirical, rather than conceptual, evaluation of governance-mechanism and literacy-intervention effectiveness; equity-focused study of adaptive and generative AI use in under-resourced and Global South institutional contexts; and longitudinal investigation of whether current patterns of GenAI use in higher education strengthen or erode graduates' higher-order cognitive and professional judgement over time.

These conclusions are interpretive rather than causal, for the reasons set out in the discussions around the trustworthiness. Future research should test the proposed framework through longitudinal, cross-institutional and Global South-inclusive designs.

Acknowledgment

The authors thank the University of Johannesburg for institutional support during the preparation of this manuscript.

Author contributions

In accordance with the CRediT (Contributor Roles) taxonomy, author contributions were as follows:

- Kurt Naicker: Conceptualisation; writing – review and editing; project correspondence (conference submission).
- Katlego Thipe: Conceptualisation; literature search and

screening; data curation (literature matrix); formal analysis (directed qualitative content analysis); writing – original draft; writing – review and editing.

Both authors approved the final version of the manuscript.

Refences

[1] S. Chakraborty, "Researching how AI can personalise learning paths, automate content creation, and provide intelligent feedback: Exploring generative AI in learning material creation and AI-powered assessment," International Journal on Science and Technology, 16(3), 1–6, 2025.

[2] J. Garzón, E. Patiño, and C. Marulanda, "Trends, benefits, and challenges of AI integration in education: A systematic review," Multimodal Technologies and Interaction, 9(6), Article 84, 2025. doi: 10.3390/mti9060084.

[3] K. Mere, "Exploring the role of generative AI and deep learning in personalized education content: A systematic literature review," AL-ISHLAH: Jurnal Pendidikan, 17(4), 2025. doi: 10.35445/alishlah.v17i4.8285.

[4] S. S. Lee and R. L. Moore, "Harnessing generative AI (GenAI) for automated feedback in higher education: A systematic review," Online Learning, 28(3), 82–104, 2024. doi: 10.24059/olj.v28i3.4593.

[5] I. M. García-López and L. Trujillo-Liñán, "Ethical and regulatory challenges of generative AI in education: A systematic review," Frontiers in Education, 10, Article 1565938, 2025. doi: 10.3389/feduc.2025.1565938.

[6] K. S. N. Rahmawati, O. A. Suciptaningsih, and A. E. Anggraini, "Generative AI in educational research: Opportunities, integrity risks, ethical challenges—A PRISMA-guided systematic review," Journal of Innovation and Research in Primary Education, 5(1), 1522–1536, 2026. doi: 10.56916/jirpe.v5i1.3106.

[7] N. Pireci Sejdiu, S. Grassini, and S. Sejdiu, "Generative AI in education: Process-aware pedagogy, assessment integrity, and institutional governance [Version 1; peer review: 1 approved, 2 approved with reservations]," Open Research Europe, 6, Article 161, 2026. doi: 10.12688/openreseurope.23343.1.

[8] M. Belkina, S. Daniel, S. Nikolic, R. Haque, S. Lyden, P. Neal, S. Grundy, and G. M. Hassan, "Generative artificial intelligence in higher education: A systematic review of empirical case studies and practical guidance," Computers and Education: Artificial Intelligence, 8, Article 100475, 2025. doi: 10.1016/j.caeai.2025.100475.

[9] X. Wang, Z. Zainuddin, and C. H. Leng, "Generative artificial intelligence in pedagogical practices: A systematic review of empirical studies (2022–2024)," Cogent Education, 12(1), Article 2485499, 2025. doi: 10.1080/2331186X.2025.2485499.

[10] S. O. Owidi, J. N. Lyanda, E. O. Okono, and E. M. Micheni, "The dark side of generative AI in higher education: Challenges, ethical concerns, and implications," East African Journal of Arts and Social Sciences, 8(3), 526–543, 2025. doi: 10.37284/eajass.8.3.3690.

[11] I. Kaşarcı, Z. Akın Demircan, G. Çeliker Ercan, and T. İnci, "Managing artificial intelligence ethics in higher education: A systematic framework for issues and policy recommendations," International Journal of Current Educational Studies, 4(2), 112–137, 2025. doi: 10.46328/ijces.223.

[12] M. I. I. Alfiras, A. Q. Emran, and A. M. Mohamed, "Ethics and governance of generative AI in education: A systematic review on responsible adoption," Discover Education, 5, Article 37, 2026. doi: 10.1007/s44217-025-01051-y.

[13] T. Tillmanns, A. Salomão Filho, S. Rudra, P. Weber, J. Dawitz, E. Wiersma, D. Dudenaite, and S. Reynolds, "Mapping tomorrow's teaching and learning spaces: A systematic review on GenAI in higher education," Trends in Higher Education, 4(1), Article 2, 2025. doi: 10.3390/higheredu4010002.

[14] M. Sharma, "The ethical considerations of using Gen AI and AI tools in academic writing in higher education: A systematic review," Kalika Journal of Multidisciplinary Research, 3(1), 82–93, 2025.

[15] M. Á. Rodríguez-Ortiz, P. C. Santana-Mancilla, and L. E. Anido-Rifón, "Machine learning and generative AI in learning analytics for higher education: A systematic review of models, trends, and challenges," Applied Sciences, 15(15), Article 8679, 2025. doi: 10.3390/app15158679.

[16] T. Chanpradit, "Generative artificial intelligence in academic writing in higher education: A systematic review," Edelweiss Applied Science and Technology, 9(4), 889–906, 2025. doi: 10.55214/25768484.v9i4.6128.

[17] R. Chinoracky and N. Stalmasekova, "Ethical problems in the use of artificial intelligence by university educators," Education Sciences, 15(10), Article 1322, 2025. doi: 10.3390/educsci15101322.

[18] G. A. Thomas, S. Abu, A. B. Ajayi, P. A. Frempong, D. Mensah, J. O. Badmus, and A. H. Etim, "Navigating generative AI in education: A narrative review of teaching practices and ethical challenges," Asian Journal of Education and Social Studies, 51(11), 507–521, 2025. doi: 10.9734/ajess/2025/v51i112630.

[19] M. S. Latif, A. S. Bhutta, N. N. Gondal, and S. Yasin, "AI-driven personalized learning in higher education: A systematic review through the lens of Bloom's taxonomy," ACADEMIA International Journal for Social Sciences, 4(4b), 277–300, 2025. doi: 10.63056/academia.4.4(b).2025.1618.

[20] H.-F. Hsieh and S. E. Shannon, "Three approaches to qualitative content analysis," Qualitative Health Research, 15(9), 1277–1288, 2005. doi: 10.1177/1049732305276687.

[21] A. B. S. Nayak and S. N. Radhika, "Generative artificial intelligence in higher education: Applications, challenges and future directions," International Journal of Science and Research, 15(5), 1440–1441, 2026. doi: 10.21275/SR26522155915.

[22] Y. Jiang, Y. Fan, and Z. Liu, "Generative AI in art education: A systematic review of research trends, tool applications, and outcomes (2019–2025)," Education Sciences, 16(1), Article 47, 2026. doi: 10.3390/educsci16010047.

[23] T. Yuensook, T. Jantakoon, and P. Limpinan, "AI-driven adaptive learning systems in higher education: A systematic review," Journal of Education and Learning, 15(2), 117–132, 2026. doi: 10.5539/jel.v15n2p117.

[24] A. Fortuna, F. Prasetya, A. D. Samala, S. Rawas, S. Criollo-C, D. Kaya, M. Raihan, W. Andriani, D. Safitri, and R. A. Nabawi, "Artificial intelligence in personalized learning: A systematic literature review," Social Sciences & Humanities Open, 12, Article 101896, 2025. doi: 10.1016/j.ssaho.2025.101896.

[25] H. Farhood, M. Nyden, A. Beheshti, and S. Muller, "Artificial intelligence-based personalised learning in education: A systematic literature review," Discover Artificial Intelligence, 5, Article 324, 2025. doi: 10.1007/s44163-025-00598-x.

[26] J. Peng and Y. Li, "Frontiers of artificial intelligence for personalized learning in higher education: A systematic review of leading articles," Applied Sciences, 15(18), Article 10096, 2025. doi: 10.3390/app151810096.

[27] X. Yuan, Z. Ding, Z. Zhang, Y. Zhou, L. Zhou, and Z. Ge, "Personalized learning resources created by generative artificial intelligence: A systematic review," In Proceedings of the 2025 11th International Conference on Education and Training Technologies (pp. 84–89). IEEE, 2025. doi: 10.1109/ICETT66247.2025.11137092.

[28] C. Coman, V. Gherheș, A. Bucs, E. Coman, V.-R. Cioca, D.-C. Bódi, and R.-M. Șerbănescu, "Navigating opportunities and challenges of generative AI in higher education: Towards responsible, equitable, and human-centered integration," Frontiers in Artificial Intelligence, 9, Article 1750978, 2026. doi: 10.3389/frai.2026.1750978.