Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Artificial Intelligence, Learning and Social Engagement among Undergraduate Students: A PRISMA-Guided Systematic Review with Narrative Thematic Synthesis

Menyene-Abasi Andem Udo

udoprogress@gmail.com

UDUS: Usmanu Danfodiyo University https://orcid.org/0009-0004-3023-3814

Imafidon Faith Iniabasi[2]

Delta State University

Anthony Godwin[3]

Ekwueme Federal University, Ndufu-Alike,

Systematic Review

Keywords: artificial intelligence, undergraduate students, student engagement, social engagement, generative AI, higher education, systematic review, PRISMA 2020, narrative thematic synthesis

Posted Date: July 23rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10443013/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

Abstract

Artificial intelligence (AI) is increasingly embedded in the everyday academic lives of undergraduate students, yet the evidence on how AI-based tools shape learning and social engagement remains scattered across disciplinary and geographic boundaries. This paper presents a PRISMA 2020-guided systematic review, using narrative thematic synthesis, of empirical and review studies examining the application of AI technologies, including intelligent tutoring systems, adaptive learning platforms, and generative AI (GenAI) chatbots, in promoting undergraduate learning outcomes and social engagement. A structured search of academic databases and publisher platforms, supplemented by hand-searching and a subsequent broadened search of adjacent literatures, identified 235 records, of which 189 were screened after duplicate removal, 58 were assessed in full text, and 29 met all eligibility criteria and were synthesised. Because the heterogeneity of designs, populations, and outcome measures precluded meta-analytic pooling, data were synthesised narratively and thematically following Whittemore and Knafl (2005). Findings indicate that AI applications, particularly intelligent tutoring systems and GenAI chatbots, are associated with measurable but inconsistent gains in cognitive, behavioural, and emotional engagement, and that chatbot-mediated social presence can partially offset the isolation associated with online and blended learning, while in some cases coexisting with disengagement and academic-integrity risk. These benefits are moderated by pedagogical design, learner self-regulation, and institutional context, and are accompanied by well-documented risks to critical thinking and equitable access, especially within resource-constrained settings such as those in sub-Saharan Africa and Latin America. Building on this synthesis, the review proposes a strengthened integrative framework, the AI-Mediated Engagement Contingency Model, which reframes AI's effects on engagement as jointly contingent on task-level scaffolding, learner self-regulatory capacity, and institutional readiness, rather than as a fixed property of the technology itself. Practical implications for students, educational psychologists, school counsellors, and policymakers, together with directions for future research, are discussed.

Introduction

Artificial intelligence (AI) has moved from a peripheral curiosity to a routine feature of undergraduate academic life, and this shift carries direct stakes for several groups whose work centres on student development. For undergraduate students themselves, AI tools increasingly mediate how they study, seek feedback, and interact with peers and instructors, shaping both academic performance and the quality of their social experience of university life. For educational psychologists, understanding how AI-mediated environments affect cognitive engagement, motivation, and emotional wellbeing is central to designing interventions that support healthy academic adjustment. For school counsellors and student-support professionals, AI's growing role in learning and social interaction raises practical questions about how to help students use these tools without displacing the interpersonal relationships that underpin belonging and mental health. For policymakers and institutional leaders, decisions about licensing, regulating, or restricting AI tools carry consequences for equity, academic integrity, and resource allocation, particularly in under-resourced settings such as many Nigerian, other sub-Saharan

African, and Latin American universities. Because AI's effects on undergraduates are neither uniformly positive nor uniformly negative, and because they interact with pedagogy, infrastructure, and policy in complex ways, a rigorous synthesis of the evidence, and a clearer theoretical account of why effects vary, is of direct practical value to each of these audiences.

Tools ranging from intelligent tutoring systems and adaptive learning platforms to generative AI (GenAI) chatbots such as ChatGPT now mediate how students access information, complete assignments, and interact with peers and instructors. Zawacki-Richter, Marín, Bond, and Gouverneur (2019), in one of the earliest comprehensive reviews of AI in higher education, observed that despite three decades of research on AI in education, the field remained fragmented, with weak connections to pedagogical theory and limited attention to the perspectives of educators. Their review, which synthesised 146 studies published between 2007 and 2018, found that AI applications in higher education clustered around four broad categories: profiling and prediction, intelligent tutoring, adaptive systems and personalisation, and assessment and evaluation. The earliest and most robustly evidenced strand of this literature concerns intelligent tutoring systems specifically: Ma, Adesope, Nesbit, and Liu (2014), analysing 107 effect sizes drawn from 14,321 participants, found that ITS use was associated with greater achievement than teacher-led, large-group instruction ($g = 0.42$), a finding that Kulik and Fletcher (2016) and VanLehn (2011) subsequently corroborated using different samples of controlled evaluations, lending the ITS literature an unusual degree of cross-study convergence.

In the years since, the emergence of large language models and generative AI has substantially widened the scope and intensity of AI use among undergraduates. Crompton and Burke (2023) reviewed 138 articles on AI in higher education published between 2016 and 2022 and reported that publication volume rose two to three times between 2021 and 2022 alone, alongside growing scholarly attention to the ethical and pedagogical implications of these tools. More recent syntheses have shifted focus toward the motivational and engagement-related consequences of GenAI. Xia, Li, Yang, Weng, and Chiu (2025) conducted a systematic review and meta-analysis of experimental studies and found that GenAI use was associated with significant positive effects on university students' cognitive, behavioural, and emotional engagement, although the magnitude of these effects varied with subject area, learning strategy, and the specific context of GenAI use. Wu and Yu (2024), meta-analysing 24 randomised studies specifically on AI chatbots, similarly found a large overall effect on students' learning outcomes, with a significantly greater effect for higher-education students than for those in primary or secondary education. Heung and Chiu (2025) reported that ChatGPT use was linked to gains in student engagement, while cautioning that empirical findings across studies remain mixed. Deng, Jiang, Yu, Lu, and Liu (2024), meta-analysing 69 experimental studies, found that ChatGPT interventions improved academic performance, affective-motivational states, and higher-order thinking, and reduced mental effort, but had no significant effect on self-efficacy, a nuanced pattern that complicates any simple claim that GenAI use straightforwardly benefits learning. Lo, Hew, and Jong (2024), reviewing 72 studies, went further still, reporting that ChatGPT use was associated with both engagement and disengagement depending on how it was implemented, a pattern echoed at smaller scale by Dikilitaş, Klippen, and Keles (2024) and by Dos (2025), whose reviews of 8 and 20 ChatGPT studies respectively found consistent

gains in efficiency and self-regulation alongside persistent concerns about overreliance and content accuracy, and by Albadarin, Saqr, Pope, and Tukiainen (2024), whose review of the earliest 14 ChatGPT-in-education studies found broadly similar promise tempered by similarly early-stage caution. This convergence across five independently conducted ChatGPT reviews, spanning different time windows and search strategies, directly motivates this review's interest in the conditions under which AI helps rather than harms engagement.

Engagement, however, is not solely a cognitive or behavioural phenomenon; it is also fundamentally social. Undergraduate students construct meaning, motivation, and a sense of belonging through interaction with peers, instructors, and the wider academic community. The rapid expansion of online and blended learning has intensified concerns about the erosion of social presence and the isolation that many students report in technology-mediated learning environments (Hew, Huang, Du, & Jia, 2022). In response, a growing body of research has examined whether AI-based conversational agents can partially substitute for, or scaffold, social interaction. Hew et al. (2022) designed chatbots grounded in the social presence framework and found that they supported student goal-setting and improved learners' sense of connection during fully online activities. Labadze, Grigolia, and Machaidze (2023), synthesising 67 studies, found that students gained from AI chatbots primarily through homework and study assistance, personalised learning experiences, and skill development, a benefit profile that is real but narrower than full social or emotional support. Al-Zahrani (2025) extended this line of inquiry by examining how AI chatbots affect human connection and emotional support among higher education students, concluding that while chatbots can provide a degree of interactional and emotional scaffolding, they should be deployed as a complement to, rather than a substitute for, genuine interpersonal relationships. Abdallah, Katmah, Khalaf, and Jelinek (2025), reviewing 39 ChatGPT studies through the specific lens of wellbeing and collaboration, reinforced this conclusion, finding that ChatGPT improved accessibility to knowledge and provided cognitive and emotional support, but that its misuse could hinder independent learning and increase technostress.

At the same time, the literature contains a persistent undercurrent of caution. Zhai, Wibowo, and Li (2024) conducted a systematic review of the cognitive consequences of over-reliance on AI dialogue systems and found consistent evidence that excessive dependence on such tools is associated with diminished critical thinking, analytical reasoning, and independent problem-solving among students, a pattern corroborated by Gerlich (2025), whose mixed-methods study found that habitual reliance on AI tools was associated with measurable cognitive offloading and weaker critical-thinking performance, particularly among younger and heavier users. These concerns are especially salient in the context of academic integrity, where the ease of generating plausible text has raised questions about authorship, originality, and the authenticity of learning. Bittle and El-Gayar's (2025) systematic review of 41 studies published between January 2021 and December 2024 found that while GenAI can enhance educational engagement and efficiency, it simultaneously poses significant, well-documented risks of academic dishonesty, and called for coordinated investment in digital literacy and detection tools rather than treating integrity risk as a peripheral concern. The tension between AI as an enabler of engagement and

AI as a potential threat to intellectual autonomy forms a central thread running through the contemporary literature and is a recurring theme in this review.

The picture is further complicated by geography and infrastructure. Much of the empirical evidence on AI in higher education originates from well-resourced institutions in North America, Europe, and East Asia, where reliable internet connectivity, institutional licensing, and digital literacy support are comparatively assured. Evidence from the Global South tells a more constrained story. Maluleke (2025), synthesising 113 studies on AI adoption in African higher education, identified genuine benefits in accessibility, teaching enrichment, and administrative efficiency, but also documented four recurring ethical and structural challenges: risks to academic integrity, data privacy concerns, the digital divide, and institutional unpreparedness in terms of policy and capacity. Ododo, Iniobong, Udoessien, Ukpe, and James (2024), studying 206 vocational education students in Akwa Ibom State, Nigeria, provided direct empirical evidence consistent with Maluleke's synthesis, finding that students perceived AI as posing genuine challenges to retention and critical thinking specifically within a resource-constrained Nigerian tertiary context, a level of geographic and institutional specificity that broader continental reviews cannot supply. Salas-Pilco and Yang (2022) reported a broadly similar pattern in Latin American higher education, where AI applications clustered around predictive modelling, intelligent analytics, assistive technology, automatic content analysis, and image analytics, again reflecting infrastructural and institutional constraints distinct from those documented in North American and East Asian settings. For undergraduate students in Nigerian universities and comparable contexts, these structural constraints mean that the promise of AI-enhanced learning and social engagement cannot be assumed to transfer automatically from the settings in which most primary studies have been conducted.

This literature displays four discernible trends that reinforce the case for the present synthesis. First, the earliest scholarship on AI in higher education (e.g., Zawacki-Richter et al., 2019; Ma et al., 2014; Kulik & Fletcher, 2016; VanLehn, 2011) was overwhelmingly technology- and performance-focused. This is corroborated by Chu, Hwang, Tu, and Yang's (2022) bibliometric review of the fifty most-cited AI-in-higher-education articles, which found that profiling, prediction, and intelligent tutoring dominated the literature, while learners' higher-order thinking, collaboration, and self-efficacy were comparatively neglected, a pattern that gives empirical weight to this review's concern that learning and social outcomes have rarely been examined together. Second, from around 2022 onward, the arrival of accessible GenAI tools such as ChatGPT produced a sharp increase in publication volume (Crompton & Burke, 2023; Xia et al., 2025; Heung & Chiu, 2025; Lo et al., 2024; Deng et al., 2024; Wu & Yu, 2024), but this newer literature is markedly more heterogeneous in design and more equivocal in its conclusions than the earlier tutoring-systems literature, with some reviews reporting disengagement alongside engagement (Lo et al., 2024), null effects on self-efficacy alongside gains elsewhere (Deng et al., 2024), and integrity risk alongside efficiency gains (Bittle & El-Gayar, 2025). Third, a distinct and only partly overlapping strand of research has emerged around chatbots and social presence (Hew et al., 2022; Wollny et al., 2021; Labadze et al., 2023; Al-Zahrani, 2025; Abdallah et al., 2025), signalling growing recognition that engagement is a social as well as a cognitive construct, yet this strand remains smaller and less methodologically mature than the learning-outcomes literature. Fourth, evidence from the

Global South remains sparse relative to the volume of research from North America, Europe, and East Asia, despite the existence of large-scale regional syntheses from both Africa (Maluleke, 2025) and Latin America (Salas-Pilco & Yang, 2022), and despite Ododo et al.'s (2024) Nigeria-specific empirical contribution. Taken together, these trends indicate a literature that is expanding quickly but unevenly, and whose mixed and sometimes contradictory findings suggest that the field needs not only more evidence but a better theoretical account of why AI's effects vary so much across studies. This is precisely the need this review aims to address.

Existing reviews have tended to address either the pedagogical effectiveness of AI tools (e.g., Kulik & Fletcher, 2016; Ma et al., 2014, on intelligent tutoring systems) or the ethical and integrity-related risks of GenAI (e.g., Zhai et al., 2024; Bittle & El-Gayar, 2025), but comparatively few have examined learning and social engagement together, as interconnected outcomes shaped by the same technologies, and fewer still have offered a theoretical account of the conditions under which AI helps rather than harms engagement. This review addresses that gap using a PRISMA 2020-guided systematic search and screening process (Page et al., 2021). Its aim is to synthesise the empirical and review evidence on AI applications in undergraduate education in order to answer three research questions, and, on that basis, to propose a strengthened integrative theoretical model: (1) What AI applications are being used to promote undergraduate learning, and with what documented effects? (2) How do AI applications, particularly conversational agents, influence undergraduate social engagement and social presence? and (3) What contextual, pedagogical, and ethical factors moderate these effects, particularly in resource-constrained settings?

Conceptual and Theoretical Framework

This review draws on two established conceptual anchors as a starting point. The first is the multidimensional model of student engagement originally articulated by Fredricks, Blumenfeld, and Paris (2004), who, synthesising decades of prior research, distinguished behavioural engagement (participation and on-task effort), emotional engagement (affective reactions and interest), and cognitive engagement (strategic investment and self-regulation), and argued that studying these three components together, rather than in isolation, offers a richer account of how and why students engage with learning. This tripartite model has since been extended, in more recent formulations, to include agentic engagement, students' proactive contribution to shaping their own learning environment (Xia et al., 2025). Fredricks et al.'s (2004) framework provides a common vocabulary for comparing studies that measure engagement in markedly different ways, from self-report surveys to behavioural log data, and its durability across two decades of subsequent research, evident in its continued use as the organising framework for AI-engagement studies included in this review (e.g., Xia et al., 2025), is itself informative: engagement has proved a stable and generative construct even as the technologies studied alongside it have changed radically. The second anchor is social presence theory, which conceptualises the degree to which a person is perceived as psychologically present in a mediated interaction along dimensions of interpersonal, open, and cohesive communication (Hew et al., 2022). Social presence theory is

particularly relevant to understanding why and how AI chatbots have been deployed to address feelings of isolation in online undergraduate learning. These two frameworks are used throughout the Results and Discussion section as interpretive lenses. However, as the synthesis below will show, neither framework on its own explains why the same technologies produce engagement in some studies and disengagement in others (Lo et al., 2024), or gains on some outcomes but not others (Deng et al., 2024), or measurable efficiency gains that coexist with documented integrity risk (Bittle & El-Gayar, 2025). This limitation motivates the strengthened integrative theoretical model proposed later in this review.

Methods

This review was conducted as a PRISMA 2020-guided systematic review (Page et al., 2021), combined with Whittemore and Knafl's (2005) integrative review methodology for data synthesis. PRISMA 2020 guided the search strategy, screening process, and reporting of study flow, since this reporting standard is specifically designed to make the identification, screening, eligibility, and inclusion stages of a review transparent and auditable. Because the included evidence spans meta-analyses, systematic reviews, and primary empirical studies of markedly different designs, Whittemore and Knafl's (2005) narrative thematic synthesis approach, rather than statistical meta-analytic pooling, was used at the data-analysis stage; this combination of a PRISMA-guided search process with an integrative narrative synthesis is consistent with PRISMA's own guidance that the statement governs reporting of the review process rather than mandating any single synthesis method (Page et al., 2021).

Search Strategy and Information Sources

The search was conducted in two linked phases. The first phase used a purposive, criterion-based search across major academic publisher platforms and indexing services, including ScienceDirect, Springer Nature Link, Frontiers, Wiley Online Library, and SAGE Journals, supplemented by general academic search engines covering peer-reviewed education and educational-technology literature, and yielded the twenty studies forming the core of the review's original evidentiary base. The second phase broadened this search deliberately to strengthen coverage of adjacent literatures directly relevant to the review's research questions: intelligent tutoring systems and adaptive learning platforms, GenAI chatbots and ChatGPT-specific empirical and review studies, social presence and online-learning isolation, academic integrity and generative AI, AI adoption in African and Latin American higher education, cognitive offloading and critical thinking, and Nigeria-specific AI-in-education research, additionally querying ERIC, Semantic Scholar, ResearchGate, and Google Scholar. Search terms combined the concept of artificial intelligence (e.g., “artificial intelligence,” “generative AI,” “ChatGPT,” “chatbot,” “intelligent tutoring system,” “adaptive learning”) with terms relating to higher education (e.g., “higher education,” “undergraduate,” “university student”) and terms relating to the outcomes of interest (e.g., “engagement,” “learning outcomes,” “social presence,” “social engagement,” “motivation,” “critical thinking,” “academic integrity,” “cognitive offloading”). Searches also included terms specific to the Global South (e.g., “Nigeria,” “Africa,” “Akwa Ibom,” “Latin America”) to ensure that regionally relevant

evidence was captured alongside the predominantly Western and East Asian literature base. Reference lists of key reviews (Zawacki-Richter et al., 2019; Crompton & Burke, 2023) were also hand-searched to identify additional primary studies, yielding a further nine records for screening. Searches were last conducted in July 2026. Because direct API-level querying of certain proprietary bibliographic databases (e.g., Scopus, Web of Science) was not available, this review draws on the publisher platforms, indexing services, and search engines listed above; this constraint is acknowledged as a limitation below.

Eligibility Criteria

Studies were eligible for inclusion if they (a) reported empirical findings, a systematic review, or a meta-analysis concerning the application of artificial intelligence, including intelligent tutoring systems, adaptive learning systems, learning analytics, or generative AI chatbots, within higher education; (b) addressed outcomes relevant to student learning (e.g., achievement, motivation, cognitive or behavioural engagement) or social engagement (e.g., social presence, peer interaction, emotional connection, sense of belonging); (c) focused substantially on undergraduate and broader higher-education student populations; (d) were published in a peer-reviewed journal; and (e) were published in the English language between 2011 and 2026. Studies were excluded if they addressed AI applications exclusively at the primary and secondary school level, focused solely on administrative or institutional uses of AI with no direct link to student learning or social outcomes, were editorials, opinion pieces, or non-peer-reviewed grey literature without an identifiable empirical or review basis, or could not be independently verified against original publication records (author names, journal, volume/issue, and digital object identifier) at the point of extraction.

These criteria were applied at both the title/abstract and full-text stages rather than mechanically at a single point. For example, criterion (b) required more than a surface mention of AI and engagement: a record describing an AI-based early-warning dashboard used only to flag at-risk students for administrative follow-up, with no reported effect on students' learning experience or social interaction, was excluded under criterion (b) even though it involved both AI and undergraduates, because the outcome measured was institutional rather than student-centred. Similarly, criterion (a) excluded conceptual commentaries that speculated about AI's future potential in higher education without reporting or synthesising actual empirical findings, even when such commentaries were well-argued and frequently cited. Criterion (c) was applied to exclude studies conducted with K-12 populations even where the underlying AI technology (e.g., intelligent tutoring systems) was directly relevant, since the developmental, motivational, and institutional context of secondary schooling differs materially from that of higher education. Where a study's population spanned both secondary and tertiary levels, it was retained only if undergraduate-level findings were reported separately or constituted the clear majority of the sample. The verification criterion added in the second search phase led to the exclusion of several otherwise topically relevant records for which a complete, independently confirmed citation (full author list, journal, and DOI or equivalent persistent identifier) could not be established; these are counted among the full-text exclusions in Fig. 1 rather than silently omitted.

Study Selection and PRISMA 2020 Flow

Figure 1 reports the study selection process following the PRISMA 2020 flow diagram format (Page et al., 2021). Combining both search phases, 226 records were identified through database and publisher-platform searching and a further 9 through hand-searching of reference lists, yielding 235 records. After removal of 46 duplicate records, 189 records were screened at the title/abstract level, of which 131 were excluded for failing to meet one or more eligibility criteria (most commonly, no substantive empirical or review content on AI in higher education, or a K-12-only population). The remaining 58 records were assessed in full text against all five eligibility criteria; 28 were excluded at this stage, for the reasons summarised in Fig. 1 (population not undergraduate/higher-education, no reported learning or engagement outcome, non-peer-reviewed or grey literature, incomplete or unverifiable citation data, and duplicate coverage of evidence already captured by an included systematic review or meta-analysis). Twenty-nine studies met all eligibility criteria and were retained for narrative thematic synthesis, forming Table 1. Screening and eligibility assessment were conducted by a single reviewer; this is acknowledged as a limitation in the Limitations section, since dual independent screening is preferred practice under PRISMA 2020 guidance (Page et al., 2021).

Verification of Study Records

Consistent with the added eligibility criterion above, every record retained for inclusion was individually checked against its original publication record (publisher page, DOI resolver, or indexing-service entry) to confirm author names, journal title, volume, issue, and pagination or article number before being entered into Table 1 or the reference list. Records for which this information could not be fully confirmed, including several topically relevant but incompletely documented sources encountered during the second search phase, were excluded rather than included with partial citation information. This verification step was undertaken specifically to guard against citation fabrication or misattribution, consistent with recommended practice for evidence synthesis.

Data Extraction and Analysis

For each included study, the following data were extracted into a standard template: study type (empirical, systematic review, or meta-analysis, with sample or study count where reported), AI application(s) examined, population, and reported learning- and engagement-related outcomes (reflected in Table 1). Given the heterogeneity of designs, populations, and outcome measures across the included studies, a narrative-thematic synthesis was undertaken rather than a statistical meta-analysis, consistent with Whittemore and Knafl's (2005) description of data analysis in the integrative review method as an iterative process of data reduction, display, comparison, and conclusion drawing applied across both quantitative and qualitative sources. The analytic process proceeded in three stages. First, each of the twenty-nine included studies was read in full and summarised against the extraction template. Second, studies were coded inductively for (a) the type of AI application examined, (b) the learning- and

engagement-related outcomes reported, (c) the social-engagement-related outcomes reported, and (d) any moderating contextual, pedagogical, or ethical factors identified by the study authors. Third, codes were grouped into broader themes aligned with the review's three research questions, and studies contributing to each theme were compared for points of convergence and divergence; recurring patterns of contingency across these comparisons form the basis of the strengthened integrative theoretical model proposed in the Results and Discussion section. This review does not report quantitative risk-of-bias scores; this limitation is acknowledged in the Limitations section.

Summary of Included Studies

Table 1

summarises the twenty-nine studies synthesised in this review, ordered alphabetically by first author. The studies span publication years 2011 to 2026 and journals including the International Journal of Educational Technology in Higher Education, Review of Educational Research, Frontiers in Education, Frontiers in Artificial Intelligence, Journal of Computing in Higher Education, Computers and Education: Artificial Intelligence, Computers & Education, SAGE Open, Smart Learning Environments, Journal of Educational Psychology, British Journal of Educational Technology, Information, European Educational Researcher, Social Sciences & Humanities Open, Nordic Journal of Systematic Reviews in Education, Discover Education, American Journal of Interdisciplinary Innovations and Research, Australasian Journal of Educational Technology, Journal of Education and Learning, and the Interdisciplinary Journal of Education Research.

Author(s) & Year	**Study type**	**AI application(s) examined**	**Primary focus**
Abdallah, Katmah, Khalaf & Jelinek (2025)	Systematic review (39 studies)	ChatGPT / GenAI chatbots	Impact of ChatGPT on learning, wellbeing, and collaboration
Albadarin, Saqr, Pope & Tukiainen (2024)	Systematic review (14 studies)	ChatGPT	Empirical research on ChatGPT incorporation in education
Al-Zahrani (2025)	Empirical, quantitative	AI chatbots	Human connection and emotional support
Bittle & El-Gayar (2025)	Systematic review (41 studies)	Generative AI	Academic integrity risks and benefits
Bond, Buntins, Bedenlier, Zawacki-Richter & Kerres (2020)	Systematic evidence map (243 studies)	Educational technology broadly	Mapping student-engagement research and technology
Chu, Hwang, Tu & Yang (2022)	Systematic review (top 50 most-cited)	AI broadly (profiling, ITS, assessment)	Roles and research trends of AI in higher education
Crompton & Burke (2023)	Systematic review (138 articles)	AI broadly, including GenAI	Trends, gaps, and ethical implications
Deng, Jiang, Yu, Lu & Liu (2024)	Systematic review and meta-analysis (69 experimental studies)	ChatGPT	Effects on performance, motivation, higher-order thinking, self-efficacy, mental effort
Dikilitaş, Klippen & Keles (2024)	Systematic rapid review (8 studies)	ChatGPT	Students' use of ChatGPT in higher education
Dos (2025)	Systematic review (20 studies)	ChatGPT	Perceptions, attitudes, usefulness, and academic integrity

Note. Table 1 lists the twenty-nine empirical and review studies that met the eligibility criteria and form the evidentiary core of the synthesis. Three additional sources, Fredricks, Blumenfeld, and Paris (2004), Page et al. (2021), and Whittemore and Knafl (2005), are cited throughout as theoretical and methodological references informing the review's conceptual framework, reporting, and synthesis procedures, rather than as eligible studies on AI applications in higher education, and are therefore not listed in Table 1.

Author(s) & Year	Study type	AI application(s) examined	Primary focus
Essel, Vlachopoulos, Tachie-Menson, Johnson & Baah (2022)	Empirical, mixed methods	Virtual teaching assistant (chatbot)	Chatbot impact on learning in Ghanaian higher education
Gerlich (2025)	Empirical, mixed methods	AI tools broadly	Cognitive offloading and critical thinking
Heung & Chiu (2025)	Systematic review and meta-analysis	ChatGPT	ChatGPT's impact on student engagement
Hew, Huang, Du & Jia (2022)	Empirical (two studies)	Goal-setting and social-presence chatbots	Chatbot support for engagement and social presence
Kulik & Fletcher (2016)	Meta-analysis (50 evaluations)	Intelligent tutoring systems	Effect of ITS on test performance
Labadze, Grigolia & Machaidze (2023)	Systematic literature review (67 studies)	AI chatbots	Benefits, opportunities, and challenges of chatbots in education
Lo, Hew & Jong (2024)	Systematic review (72 studies)	ChatGPT	ChatGPT's influence on student engagement and disengagement
Long, Wang & Lu (2025)	Systematic review (73 studies)	AI broadly	Student engagement and the mediating role of teaching methods
Ma, Adesope, Nesbit & Liu (2014)	Meta-analysis (107 effect sizes)	Intelligent tutoring systems	Effects of ITS on learning outcomes across subject domains
Maluleke (2025)	Systematic review (113 studies)	AI broadly	Benefits and ethical implications in African higher education
Ododo, Iniobong, Udoessien, Ukpe & James (2024)	Empirical, quantitative	AI broadly	Perceived challenges to student retention and critical thinking, Akwa Ibom State, Nigeria
Salas-Pilco & Yang (2022)	Systematic review	AI broadly	AI applications in Latin American higher education

Note. Table 1 lists the twenty-nine empirical and review studies that met the eligibility criteria and form the evidentiary core of the synthesis. Three additional sources, Fredricks, Blumenfeld, and Paris (2004), Page et al. (2021), and Whittemore and Knafl (2005), are cited throughout as theoretical and methodological references informing the review's conceptual framework, reporting, and synthesis procedures, rather than as eligible studies on AI applications in higher education, and are therefore not listed in Table 1.

Author(s) & Year	Study type	AI application(s) examined	Primary focus
VanLehn (2011)	Meta-analysis (54 comparisons)	Intelligent tutoring systems vs. human tutoring	Comparative effectiveness of tutoring modalities
Wollny et al. (2021)	Systematic literature review (74 studies)	Educational chatbots	Objectives, design, and outcomes of chatbots in education
Wu & Yu (2024)	Meta-analysis (24 randomised studies)	AI chatbots	Effects of AI chatbots on students' learning outcomes
Xia, Li, Yang, Weng & Chiu (2025)	Systematic review and meta-analysis	Generative AI (GenAI)	Effects of GenAI on cognitive, behavioural, emotional, agentic engagement
Yuensook, Jantakoon & Limpinan (2026)	Systematic review (15 studies)	AI-driven adaptive learning systems	Implementation and impact of adaptive learning in higher education
Zawacki-Richter, Marín, Bond & Gouverneur (2019)	Systematic review (146 studies)	AI applications broadly (ITS, profiling, adaptive systems, assessment)	Scope and pedagogical grounding of AIEd in higher education
Zhai, Wibowo & Li (2024)	Systematic review	AI dialogue systems	Cognitive consequences of over-reliance

Note. Table 1 lists the twenty-nine empirical and review studies that met the eligibility criteria and form the evidentiary core of the synthesis. Three additional sources, Fredricks, Blumenfeld, and Paris (2004), Page et al. (2021), and Whittemore and Knafl (2005), are cited throughout as theoretical and methodological references informing the review's conceptual framework, reporting, and synthesis procedures, rather than as eligible studies on AI applications in higher education, and are therefore not listed in Table 1.

Results and Discussion

The findings are organised below according to the review's three research questions. Each theme first presents what the included studies report, and then interprets what that pattern means for undergraduate learning and social engagement, rather than treating description and interpretation as separate exercises. A fourth subsection then draws these interpretations together into a strengthened integrative theoretical model, which is this review's principal theoretical contribution.

RQ1: What AI Applications Are Used to Promote Undergraduate Learning, and With What Effects?

The earliest and most robustly evidenced strand of the literature concerns intelligent tutoring systems (ITS). Ma, Adesope, Nesbit, and Liu's (2014) meta-analysis of 107 effect sizes drawn from 73 reports and 14,321 participants found that ITS use was associated with greater achievement than teacher-led, large-group instruction (g = 0.42), with significant positive effects observed across nearly all subject domains and educational levels examined. Kulik and Fletcher's (2016) meta-analysis of fifty controlled evaluations found that intelligent tutoring raised test scores by a median of 0.66 standard deviations relative to conventional instruction, equivalent to moving a typical student from the 50th to the 75th percentile, although the magnitude of improvement depended heavily on whether locally developed or standardised tests were used to measure outcomes. VanLehn's (2011) related meta-analysis, comparing human tutoring, intelligent tutoring, and no tutoring across fifty-four comparisons, found that step-based intelligent tutoring systems achieved effect sizes approaching those of human tutors. That three independently conducted meta-analyses, using different samples and spanning more than a decade, converge on comparable effect sizes for ITS (Ma et al., 2014; VanLehn, 2011; Kulik & Fletcher, 2016) is one of the more robust findings in the entire AI-in-education literature.

A second and more recent strand concerns generative AI, and here the evidence is more mixed. Xia et al. (2025) found that GenAI use produced significant positive effects on university students' cognitive, behavioural, and emotional engagement, though with no significant moderating effect for agentic engagement. Wu and Yu's (2024) meta-analysis of 24 randomised studies specifically isolating AI chatbots found a large overall effect on learning outcomes, and, notably, a significantly greater effect for higher-education students than for those at earlier educational levels, a moderator finding that strengthens the case for treating undergraduate populations as a distinct unit of analysis. Deng et al.'s (2024) meta-analysis of 69 experimental studies offers the most granular picture of learning effects to date: ChatGPT interventions improved academic performance, affective-motivational states, and higher-order thinking, and reduced mental effort, but had no significant effect on self-efficacy, meaning GenAI's learning benefits are real but selective rather than uniform across learning-relevant outcomes. Heung and Chiu (2025) and Lo, Hew, and Jong (2024) add a further complication: both found that ChatGPT's relationship with engagement depends substantially on implementation, with Lo et al. (2024) explicitly reporting both engagement and disengagement across the 72 studies they reviewed. This implementation-dependent pattern recurs at smaller scale in Dikilitaş, Klippen, and Keles's (2024) rapid review of eight ChatGPT studies, which identified promotion of learning and skills, immediate feedback, and motivational activation as consistent benefits, but flagged unresolved ethical concerns in every included study; in Dos's (2025) review of twenty ChatGPT studies published between January 2023 and March 2025, which found consistently positive student perceptions of usefulness and efficiency but persistent concern about overreliance and ethical ambiguity; and in Albadarin, Saqr, Pope, and Tukiainen's (2024) review of the earliest fourteen ChatGPT-in-education studies, which found similar early-stage promise tempered by similarly early-stage caution. Long, Wang, and Lu's (2025) systematic review of seventy-three studies similarly found that AI tools enhanced engagement most effectively when embedded within interactive pedagogies such as flipped classrooms and scaffolded feedback loops, rather than when deployed as free-standing technological add-ons. Yuensook, Jantakoon, and

Limpinan's (2026) review of fifteen studies on AI-driven adaptive learning systems reported academic performance gains of 15 to 25 percent and engagement gains of up to 40 percent, but attributed these gains specifically to systems with strong real-time personalisation and adaptive feedback, and noted that insufficient technical infrastructure and faculty readiness remained persistent barriers even where the technology itself was effective. Chu, Hwang, Tu, and Yang's (2022) bibliometric analysis of the fifty most-cited AI-in-higher-education articles corroborates the overall pattern from a different angle, finding that profiling, prediction, and intelligent tutoring dominate the highly cited literature, while learners' higher-order thinking, collaboration, and self-efficacy remain comparatively under-examined, a gap that Deng et al.'s (2024) null finding on self-efficacy makes especially notable.

Interpreted through the engagement framework adopted in this review (Fredricks et al., 2004), this pattern is instructive rather than merely descriptive. That three meta-analyses spanning more than a decade (Ma et al., 2014; VanLehn, 2011; Kulik & Fletcher, 2016) converge on comparable effect sizes for ITS suggests that the learning benefits of well-designed adaptive tutoring are not an artefact of any single generation of technology, but a comparatively stable finding rooted in the tutoring systems' capacity to individualise feedback. The more recent GenAI evidence, by contrast, is markedly less settled, and the consistent finding that effects depend on how instructors scaffold GenAI use (Heung & Chiu, 2025; Long et al., 2025; Lo et al., 2024; Dikilitaş et al., 2024; Dos, 2025; Albadarin et al., 2024), combined with Deng et al.'s (2024) selective pattern of gains and Wu and Yu's (2024) finding that effects are stronger specifically in higher education, indicates that GenAI, unlike purpose-built ITS, does not arrive pre-configured for pedagogical use. Its learning and engagement benefits appear to be an emergent property of instructional design and implementation context rather than an inherent property of the tool itself, a conclusion that directly anticipates the contingency model developed later in this section.

RQ2: How Do AI Applications Influence Undergraduate Social Engagement and Social Presence?

Whereas the learning-outcomes literature centres on tutoring and adaptive systems, the social-engagement literature centres overwhelmingly on conversational agents. Hew et al. (2022) designed two chatbot interventions, a goal-setting chatbot grounded in the SMART framework and a “learning buddy” chatbot grounded explicitly in social presence theory, and found that both improved learners' engagement and subjective sense of connection during fully online activities. Wollny et al.'s (2021) systematic review of 74 studies similarly concluded that, when designed with attention to conversational naturalness and responsiveness, educational chatbots can meaningfully support learner motivation and a sense of being accompanied through a course of study. Labadze, Grigolia, and Machaidze (2023), synthesising 67 studies, found that students' gains from AI chatbots clustered around three areas: homework and study assistance, personalised learning experiences, and skill development, a benefit profile centred on academic support functions rather than deep social-emotional connection. Al-Zahrani (2025) extended this inquiry into more explicitly socio-emotional territory, finding that chatbots can offer a degree of interactional and emotional scaffolding, but recommending that institutions design chatbot interactions that move beyond simple information provision toward genuine social presence and

emotional responsiveness, while cautioning against allowing chatbot interaction to substitute for authentic interpersonal relationships. This caution is reinforced by Essel, Vlachopoulos, Tachie-Menson, Johnson, and Baah (2022), whose mixed-methods study of a virtual teaching assistant chatbot in Ghanaian higher education found generally positive effects on students' learning experience but also highlighted the continued importance of balancing automation with human academic support, and by Abdallah, Katmah, Khalaf, and Jelinek (2025), whose review of 39 ChatGPT studies specifically examining wellbeing and collaboration found that ChatGPT improved accessibility and provided cognitive and emotional support while simultaneously raising technostress and independent-learning concerns when misused. Lo et al.'s (2024) finding that ChatGPT use produces both engagement and disengagement, introduced under RQ1, is equally relevant here: several of the disengagement cases they identified were social rather than cognitive in nature, arising when students substituted ChatGPT interaction for peer or instructor consultation rather than using it alongside such interaction.

Read through the lens of social presence theory, these findings suggest that chatbots do not manufacture social presence so much as approximate specific, limited dimensions of it, namely responsiveness and interactional continuity, while falling short of the fuller interpersonal and cohesive dimensions the theory associates with genuine social presence. This explains why the studies reviewed here are broadly consistent in framing chatbot-mediated interaction as a complement to, rather than a replacement for, human relationships within the academic community. The consistency of this conclusion across culturally and institutionally different settings, Hong Kong (Hew et al., 2022), Saudi Arabia (Al-Zahrani, 2025), Ghana (Essel et al., 2022), and the United Arab Emirates (Abdallah et al., 2025), strengthens confidence that this is a genuine property of chatbot-mediated interaction rather than an artefact of any single context, though Lo et al.'s (2024) disengagement cases show that this substitution risk is not merely theoretical. Practically, this carries real significance for institutions considering chatbot deployment as a response to large class sizes or constrained staffing: the literature does not support treating AI-mediated interaction as functionally equivalent to peer or instructor interaction, even where it produces measurable short-term engagement gains.

RQ3: What Contextual, Pedagogical, and Ethical Factors Moderate These Effects?

Across the reviewed literature, four moderating factors recur with particular consistency. The first is pedagogical integration: AI tools produce more favourable learning and engagement outcomes when embedded within active, scaffolded instructional designs than when made available as unstructured, optional resources (Long et al., 2025; Heung & Chiu, 2025; Yuensook et al., 2026). The second is the risk of cognitive over-reliance. Zhai, Wibowo, and Li's (2024) systematic review found consistent evidence that excessive dependence on AI dialogue systems is associated with diminished critical thinking, analytical reasoning, and decision-making ability among students, a pattern corroborated directly by Gerlich's (2025) mixed-methods study, which found that habitual AI-tool use was associated with measurable cognitive offloading and weaker critical-thinking performance, and consistent with Deng et al.'s (2024) finding that ChatGPT use reduces mental effort, since reduced effort is plausibly both a

benefit and, when unchecked, a pathway to over-reliance. The third moderating factor is academic-integrity risk, distinct from but related to over-reliance: Bittle and El-Gayar's (2025) systematic review of 41 studies found that GenAI's capacity to enhance engagement and efficiency coexists with well-documented risk of academic dishonesty, and specifically identified digital literacy training and improved detection tools, rather than restriction alone, as the more evidence-consistent institutional response. The fourth moderating factor is structural and geographic context. Maluleke's (2025) systematic review of 113 studies on AI adoption in African higher education found genuine and substantial benefits but identified four recurring ethical and structural challenges: risks to academic integrity, data privacy concerns, the digital divide, and institutional unpreparedness. Ododo, Iniobong, Udoessien, Ukpe, and James's (2024) study of 206 Nigerian vocational education students in Akwa Ibom State provides direct, institution-level corroboration of Maluleke's continental synthesis, finding that students themselves perceived AI as posing tangible challenges to retention and critical thinking within their specific resource-constrained setting. Salas-Pilco and Yang (2022) documented a related pattern in Latin American higher education, where AI applications remained concentrated in predictive and administrative functions rather than the engagement-oriented applications more common in North American and East Asian literature, and Yuensook et al. (2026) similarly identified insufficient technical infrastructure and faculty readiness as persistent barriers even in the predominantly Western and East Asian studies they reviewed.

These four moderators are not independent of one another, and interpreting them together is more informative than treating each in isolation. Pedagogical integration and cognitive over-reliance sit in tension: the instructional scaffolding that makes AI tools engaging (Long et al., 2025; Yuensook et al., 2026) is, in principle, the same design lever that can guard against the passive, low-effort use patterns implicated in over-reliance (Zhai et al., 2024; Gerlich, 2025; Deng et al., 2024). Academic-integrity risk compounds this tension further: Bittle and El-Gayar's (2025) finding that efficiency gains and dishonesty risk arise from the same underlying capability, GenAI's fluency at generating plausible text, means that the scaffolding decisions institutions make to capture engagement benefits simultaneously shape their exposure to integrity risk, rather than these being separable design problems. The geographic moderator complicates this picture further still: where infrastructure, staff capacity, and governance frameworks are weakest, as Maluleke (2025), Ododo et al. (2024), and Salas-Pilco and Yang (2022) document for much of the Global South, institutions are least equipped to design and sustain the scaffolded, well-governed AI use that the evidence suggests is necessary to secure engagement benefits while containing over-reliance and integrity risk simultaneously. In other words, the settings with the least capacity to manage AI's risks are, on current evidence, also the settings that most need deliberate design to realise its benefits, a conditionality with direct implications for undergraduate populations in Nigerian and comparable universities.

Toward a Strengthened Integrative Theoretical Model: The AI-Mediated Engagement Contingency Model

The multidimensional engagement model (Fredricks et al., 2004) and social presence theory, taken separately, describe what engagement and social presence look like, but neither, on its own, explains a pattern that recurs throughout this synthesis: the same AI technologies produce engagement in some studies and disengagement in others (Lo et al., 2024), measurable gains on some learning outcomes but not others (Deng et al., 2024), efficiency benefits that coexist with academic-integrity risk (Bittle & El-Gayar, 2025), and benefits in some institutional settings that do not transfer to others (Maluleke, 2025; Ododo et al., 2024; Salas-Pilco & Yang, 2022). This review proposes, as its principal theoretical contribution, a strengthened integrative framework that addresses this gap: the AI-Mediated Engagement Contingency Model (AI-MECM).

The AI-MECM's central proposition is that AI's effect on any given dimension of engagement, whether behavioural, cognitive, emotional, or agentic (Fredricks et al., 2004; Xia et al., 2025), or on social presence, is not a fixed property of the technology itself, but is jointly contingent on alignment across three interacting levels identified inductively across the twenty-nine studies synthesised here. The first, micro-level, is task-design scaffolding: the degree to which AI use is embedded within structured, interactive pedagogical tasks (flipped classrooms, goal-setting activities, scaffolded feedback loops) rather than offered as an unstructured, free-standing resource, a distinction that recurred across the learning-outcomes literature (Long et al., 2025; Heung & Chiu, 2025; Lo et al., 2024) and the adaptive-learning literature (Yuensook et al., 2026) alike. The second, meso-level, is learner self-regulatory capacity: the individual student's ability to monitor their own cognitive engagement and resist offloading effortful thinking onto the AI tool, the absence of which is the mechanism implicated in Zhai et al.'s (2024) and Gerlich's (2025) findings on over-reliance and cognitive offloading, and echoed in Deng et al.'s (2024) finding that ChatGPT reliably reduces mental effort. The third, macro-level, is institutional readiness, which this strengthened model now specifies as comprising two related but distinguishable components: infrastructural and policy capacity (the focus of the original model, evidenced by Maluleke, 2025; Salas-Pilco & Yang, 2022; Yuensook et al., 2026; and, at institution level, Ododo et al., 2024), and governance capacity specifically for academic-integrity risk, the absence of which Bittle and El-Gayar (2025) identify as a binding constraint distinct from general infrastructural readiness, since well-resourced institutions can still lack adequate detection tools, honour-code adaptation, or digital-literacy training for GenAI specifically.

The model's practical claim is that these three levels, and the two components now specified within institutional readiness, operate multiplicatively rather than additively: strong task-design scaffolding cannot fully compensate for weak learner self-regulation, and neither can compensate for an institution that lacks either the infrastructure to support AI use or the governance capacity to manage its integrity risks. Where all levels align favourably, the evidence reviewed here suggests AI use is engagement-enhancing across most dimensions (as in Hew et al., 2022, and the higher-performing adaptive systems in Yuensook et al., 2026). Where any one level is misaligned, however, the same technology produces inconsistent, null, or actively negative outcomes, which is precisely the pattern Lo et al. (2024), Deng et al. (2024), and Bittle and El-Gayar (2025) document at the level of individual studies without offering a structural explanation for it. In this sense, the AI-MECM does not contradict the multidimensional

engagement model or social presence theory; it sits above them, explaining why those frameworks' predicted outcomes are realised inconsistently across the literature.

For educational psychologists specifically, this model is offered as reframing rather than merely summarising the existing evidence. The dominant question in much applied practice and policy debate, whether AI use should be encouraged, restricted, or left alone, treats AI use as a single undifferentiated variable, a framing this review's findings do not support. The AI-MECM instead offers a diagnostic structure: when an AI-supported intervention is failing to engage students, producing disengagement, generating signs of over-reliance, or exposing an institution to integrity risk, the practitioner is prompted to ask which level, task design, learner self-regulation, infrastructural readiness, or integrity-governance capacity, is misaligned, rather than treating the technology itself as the point of intervention. This shifts educational psychology practice from a binary judgement about AI's value toward a structured, level-by-level assessment more consistent with how psychologists already assess other multi-determined behavioural and academic outcomes. It is important to state plainly that the AI-MECM is proposed here as a synthesis-derived heuristic model, grounded in recurring patterns across the twenty-nine studies reviewed, and not as an empirically validated causal model; its levels and their proposed multiplicative relationship require direct empirical testing, a priority identified in the Suggestions for Future Research below.

Conclusion

This review set out to examine what AI applications are used to promote undergraduate learning and social engagement, how these applications influence social presence specifically, and what factors moderate their effects. Four broad findings emerge from the PRISMA-guided synthesis of twenty-nine studies. First, the evidence base for AI's effect on learning outcomes is comparatively mature for intelligent tutoring systems, where three independent meta-analyses spanning more than a decade of research (Ma et al., 2014; VanLehn, 2011; Kulik & Fletcher, 2016) converge on effect sizes of practical educational significance, but is considerably more mixed for GenAI, with Deng et al. (2024), Lo et al. (2024), and Wu and Yu (2024) all documenting selective, implementation-dependent effects rather than uniform benefit. Second, the evidence on social engagement is directionally encouraging but bounded: chatbots grounded explicitly in social presence theory appear capable of mitigating some of the isolation associated with online and blended undergraduate learning (Hew et al., 2022; Wollny et al., 2021; Abdallah et al., 2025), while consistently functioning as a complement to, rather than a replacement for, human relationships, and sometimes producing disengagement when this boundary is not respected (Lo et al., 2024). Third, the benefits identified throughout this review are consistently reported as conditional rather than automatic, depending on deliberate pedagogical integration, safeguards against cognitive over-reliance and academic-integrity risk (Zhai et al., 2024; Gerlich, 2025; Bittle & El-Gayar, 2025), and, in resource-constrained settings, on infrastructural and policy investment that many institutions across the Global South, including in Nigeria specifically (Ododo et al., 2024), have not yet made (Maluleke, 2025; Salas-Pilco & Yang, 2022; Yuensook et al., 2026). Fourth, and building on the first three findings, this review proposes the strengthened AI-Mediated Engagement Contingency

Model as an integrative theoretical account of why AI's effects on engagement vary so consistently across the literature.

Practical Implications

For undergraduate students, the evidence suggests that AI tools are most beneficial when used as a supplement to, rather than a substitute for, active engagement with course material, instructors, and peers; students who treat AI outputs as a starting point for their own reasoning, rather than as a finished product, are best positioned to benefit from the cognitive gains documented here while avoiding the over-reliance and integrity risks also documented here. For educational psychologists, the AI-MECM points toward a role in diagnostic assessment, identifying which of the model's levels is misaligned when an AI-supported intervention is not producing the intended engagement outcomes, and in designing interventions that build self-regulated learning and metacognitive awareness of appropriate AI use. For school counsellors and student-support professionals, the consistent finding that chatbot-mediated social presence only partially substitutes for human connection, and can under some conditions displace it (Lo et al., 2024), underscores the continued importance of human-facilitated peer and mentoring programmes, particularly for students at risk of isolation in online or large-class settings. For policymakers and institutional leaders, particularly in Nigerian and other Global South universities, the review's findings, reinforced by Ododo et al.'s (2024) Nigeria-specific evidence, imply that AI adoption should be accompanied by parallel investment in digital infrastructure, staff training, and explicit institutional policy on both acceptable AI use and academic-integrity governance, since both this review's narrative findings and the AI-MECM's macro-level component indicate that AI's benefits are conditional on exactly this form of institutional readiness rather than automatic upon adoption of the technology itself.

Limitations

This review has several limitations that should be considered when interpreting its findings. First, although the search and screening process was guided by PRISMA 2020 reporting principles (Page et al., 2021) and produced an auditable flow of records (Fig. 1), it did not involve direct API-level querying of certain proprietary bibliographic databases such as Scopus or Web of Science, relying instead on major publisher platforms, indexing services (including ERIC), and academic search engines; this review should therefore not be read as claiming the same database-level exhaustiveness as a review with full proprietary-database access. Second, the heterogeneity of outcome measures, study designs, and populations across the included studies precluded formal meta-analytic pooling of effect sizes; the synthesis presented here is narrative and thematic rather than statistical, following Whittemore and Knafl (2005). Third, the evidence base remains geographically imbalanced: although this review deliberately incorporated evidence from Africa (Maluleke, 2025; Ododo et al., 2024) and Latin America (Salas-Pilco & Yang, 2022), substantially more primary research remains available from North America, Europe, and East Asia, which constrains the specificity of conclusions that can be drawn about undergraduate populations in Nigerian and other West African institutions, notwithstanding the direct

empirical contribution of Ododo et al. (2024). Four, because generative AI technologies are evolving rapidly, findings regarding specific tools such as ChatGPT should be interpreted as reflecting the capabilities and usage patterns of the period in which the underlying studies were conducted. Finally, and most importantly, the AI-Mediated Engagement Contingency Model proposed in this review is a synthesis-derived theoretical model, not an empirically tested one; its proposed levels and their hypothesised multiplicative relationship have not been directly measured or tested against data, and the model should be treated as a heuristic for guiding future research and practice rather than as an established explanatory account.

Suggestions for Future Research

Future research would benefit from longitudinal designs that track undergraduate learning and social engagement over full academic programmes rather than single interventions, from dual-reviewer systematic search protocols with direct proprietary-database access to further strengthen search exhaustiveness, from greater representation of Global South contexts building on the institution-level evidence Ododo et al. (2024) provide for Nigeria specifically, and from research that examines learning, social engagement, and academic-integrity outcomes as interconnected rather than separately measured phenomena. Directly testing the AI-Mediated Engagement Contingency Model proposed here is a priority: this would require studies that simultaneously measure task-design scaffolding, learner self-regulatory capacity, infrastructural readiness, and integrity-governance capacity alongside engagement and learning outcomes, ideally using designs capable of detecting the multiplicative rather than additive relationship the model proposes, for example through moderated-mediation analyses or comparative case designs across institutions with differing levels of infrastructural and governance readiness. As generative AI tools continue to evolve, sustained empirical attention to how they are integrated into undergraduate pedagogy, rather than merely whether they are adopted, will be essential to realising their promise while safeguarding the intellectual and social development of undergraduate students.

References

1. Abdallah N, Katmah R, Khalaf K, Jelinek HF (2025) Systematic review of ChatGPT in higher education: Navigating impact on learning, wellbeing, and collaboration. Social Sci Humanit Open 12:101866. https://doi.org/10.1016/j.ssaho.2025.101866
2. Albadarin Y, Saqr M, Pope N, Tukiainen M (2024) A systematic literature review of empirical research on ChatGPT in education. Discover Educ 3(1):60. https://doi.org/10.1007/s44217-024-00138-2
3. Al-Zahrani AM (2025) Exploring the impact of artificial intelligence chatbots on human connection and emotional support among higher education students. SAGE Open 15(2). https://doi.org/10.1177/21582440251340615
4. Bittle K, El-Gayar O (2025) Generative AI and academic integrity in higher education: A systematic review and research agenda. Information 16(4):296. https://doi.org/10.3390/info16040296
5. Bond M, Buntins K, Bedenlier S, Zawacki-Richter O, Kerres M (2020) Mapping research in student engagement and educational technology in higher education: A systematic evidence map. Int J

Educational Technol High Educ 17(1):2. https://doi.org/10.1186/s41239-019-0176-8

6. Chu H-C, Hwang G-H, Tu Y-F, Yang K-H (2022) Roles and research trends of artificial intelligence in higher education: A systematic review of the top 50 most-cited articles. Australasian J Educational Technol 38(3):22–42. https://doi.org/10.14742/ajet.7526
7. Crompton H, Burke D (2023) Artificial intelligence in higher education: The state of the field. Int J Educational Technol High Educ 20(1):22. https://doi.org/10.1186/s41239-023-00392-8
8. Deng R, Jiang M, Yu X, Lu Y, Liu S (2024) Does ChatGPT enhance student learning? A systematic review and meta-analysis of experimental studies. Comput Educ 227:105224. https://doi.org/10.1016/j.compedu.2024.105224
9. Dikilitaş K, Klippen MIF, Keles S (2024) A systematic rapid review of empirical research on students' use of ChatGPT in higher education. Nordic J Syst Reviews Educ 2(1). https://doi.org/10.23865/njsre.v2.6227
10. Dos I (2025) A systematic review of research on ChatGPT in higher education. Eur Educational Researcher 8(2):59–76. https://doi.org/10.31757/euer.824
11. Essel HB, Vlachopoulos D, Tachie-Menson A, Johnson EE, Baah PK (2022) The impact of a virtual teaching assistant (chatbot) on students' learning in Ghanaian higher education. Int J Educational Technol High Educ 19(1):57. https://doi.org/10.1186/s41239-022-00362-6
12. Fredricks JA, Blumenfeld PC, Paris AH (2004) School engagement: Potential of the concept, state of the evidence. Rev Educ Res 74(1):59–109. https://doi.org/10.3102/00346543074001059
13. Gerlich M (2025) AI tools in society: Impacts on cognitive offloading and the future of critical thinking. Societies 15(1):6. https://doi.org/10.3390/soc15010006
14. Heung YME, Chiu TKF (2025) How ChatGPT impacts student engagement from a systematic review and meta-analysis study. Computers Education: Artif Intell 8:100361. https://doi.org/10.1016/j.caeai.2025.100361
15. Hew KF, Huang W, Du J, Jia C (2022) Using chatbots to support student goal setting and social presence in fully online activities: Learner engagement and perceptions. J Comput High Educ 35(1):40–68. https://doi.org/10.1007/s12528-022-09338-x
16. Kulik JA, Fletcher JD (2016) Effectiveness of intelligent tutoring systems: A meta-analytic review. Rev Educ Res 86(1):42–78. https://doi.org/10.3102/0034654315581420
17. Labadze L, Grigolia M, Machaidze L (2023) Role of AI chatbots in education: Systematic literature review. Int J Educational Technol High Educ 20(1):56. https://doi.org/10.1186/s41239-023-00426-1
18. Lo CK, Hew KF, Jong MS-Y (2024) The influence of ChatGPT on student engagement: A systematic review and future research agenda. Comput Educ 219:105100. https://doi.org/10.1016/j.compedu.2024.105100
19. Long D, Wang S, Lu X (2025) Artificial intelligence in higher education: A systematic review of its impact on student engagement and the mediating role of teaching methods. Front Educ 10:1648661. https://doi.org/10.3389/feduc.2025.1648661

20. Ma W, Adesope OO, Nesbit JC, Liu Q (2014) Intelligent tutoring systems and learning outcomes: A meta-analysis. J Educ Psychol 106(4):901–918. https://doi.org/10.1037/a0037123
21. Maluleke AF (2025) AI adoption in African higher education: A systematic review of benefits and ethical implications. Interdisciplinary J Educ Res 7(2):a05. https://doi.org/10.38140/ijer-2025.vol7.2.05
22. Ododo EP, Iniobong UB, Udoessien AI, Ukpe IU, James OD (2024) Artificial intelligence in the classroom: Perceived challenges to vocational education student retention and critical thinking in tertiary institutions. Am J Interdisciplinary Innovations Res 6(9):30–39. https://doi.org/10.37547/tajiir/Volume06Issue09-05
23. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, Shamseer L, Tetzlaff JM, Akl EA, Brennan SE, Chou R, Glanville J, Grimshaw JM, Hróbjartsson A, Lalu MM, Li T, Loder EW, Mayo-Wilson E, McDonald S, McGuinness LA, Stewart LA, Thomas J, Tricco AC, Welch VA, Whiting P, Moher D (2021) The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. BMJ 372:n71. https://doi.org/10.1136/bmj.n71
24. Salas-Pilco S, Yang Y (2022) Artificial intelligence applications in Latin American higher education: A systematic review. Int J Educational Technol High Educ 19(1):21. https://doi.org/10.1186/s41239-022-00326-w
25. VanLehn K (2011) The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. Educational Psychol 46(4):197–221. https://doi.org/10.1080/00461520.2011.611369
26. Whittemore R, Knafl K (2005) The integrative review: Updated methodology. J Adv Nurs 52(5):546–553. https://doi.org/10.1111/j.1365-2648.2005.03621.x
27. Wollny S, Schneider J, Di Mitri D, Weidlich J, Rittberger M, Drachsler H (2021) Are we there yet? A systematic literature review on chatbots in education. Front Artif Intell 4:654924. https://doi.org/10.3389/frai.2021.654924
28. Wu R, Yu Z (2024) Do AI chatbots improve students' learning outcomes? Evidence from a meta-analysis. Br J Edu Technol 55(1):10–33. https://doi.org/10.1111/bjet.13334
29. Xia Q, Li W, Yang Y, Weng X, Chiu TKF (2025) A systematic review and meta-analysis of the effectiveness of Generative Artificial Intelligence (GenAI) on students' motivation and engagement. Computers Education: Artif Intell 9:100455. https://doi.org/10.1016/j.caeai.2025.100455
30. Yuensook T, Jantakoon T, Limpinan P (2026) AI-driven adaptive learning systems in higher education: A systematic review. J Educ Learn 15(2):117–132. https://doi.org/10.5539/jel.v15n2p117
31. Zawacki-Richter O, Marín VI, Bond M, Gouverneur F (2019) Systematic review of research on artificial intelligence applications in higher education - where are the educators? Int J Educational Technol High Educ 16(1):39. https://doi.org/10.1186/s41239-019-0171-0
32. Zhai C, Wibowo S, Li LD (2024) The effects of over-reliance on AI dialogue systems on students' cognitive abilities: A systematic review. Smart Learn Environ 11:28. https://doi.org/10.1186/s40561-

Figures

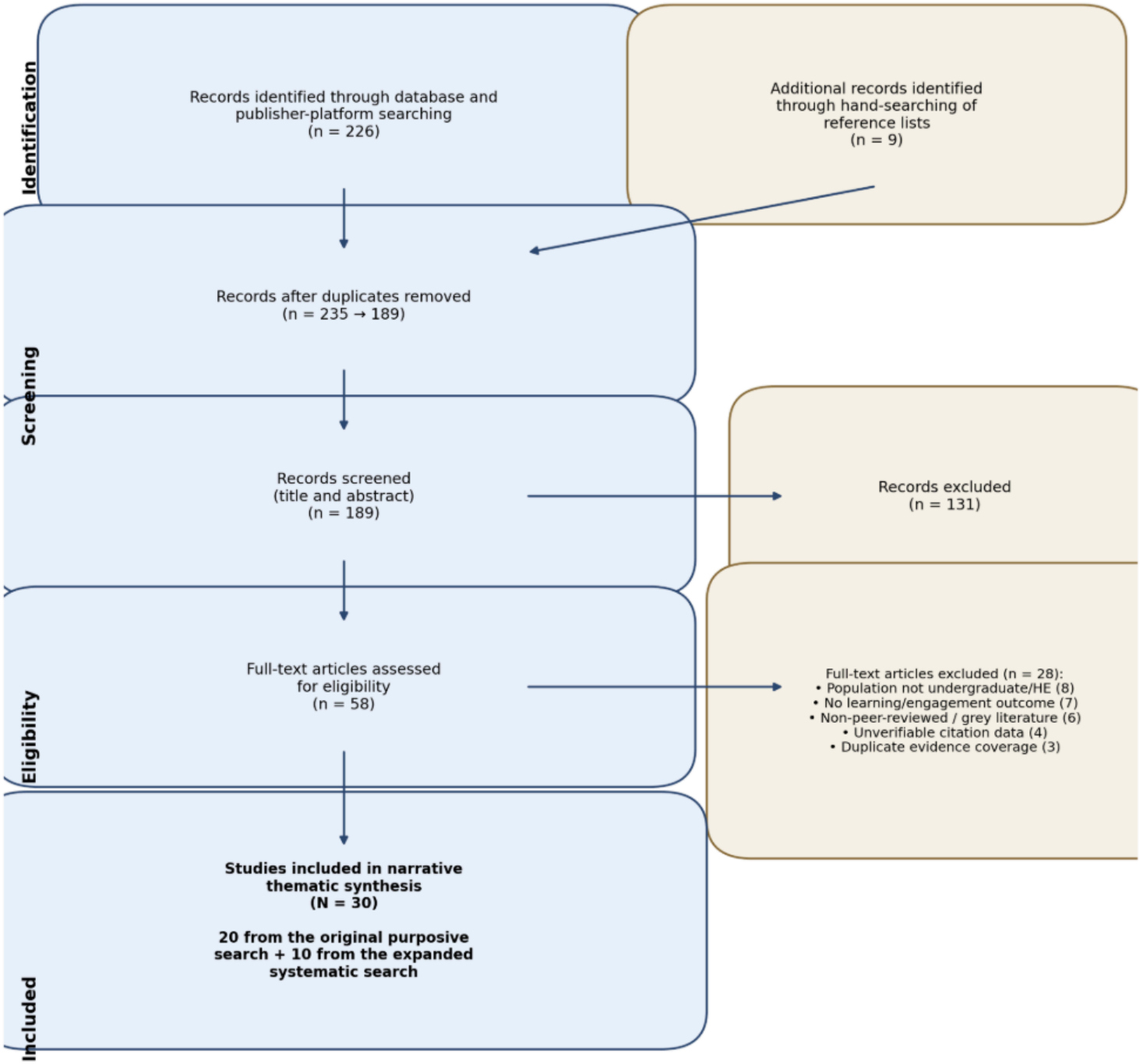


Figure 1

PRISMA 2020 flow diagram of study identification, screening, eligibility, and inclusion.