Generative Artificial Intelligence in Higher Education: A Systematic Review of Tools, Learning Outcomes, and Factors Influencing Integration

Richeno Alberto Matimbe*, Ezrone Robert Mkandawire, Doris Wangui Muturi, Bako Abdalla Hama Salih

Department of International and Comparative Education, Zhejiang Normal University, Jinhua, Zhejiang, China

***Corresponding Author**

DOI: https://doi.org/10.47772/IJRISS.2026.100900633

Received: 19 September 2026; Accepted: 05 October 2026; Published: 19 October 2026

ABSTRACT

Generative Artificial Intelligence (GenAI) is reshaping higher education by expanding opportunities for academic writing, feedback, tutoring, research support, and personalised learning. This systematic review synthesised empirical evidence on GenAI use in higher education, focusing on the tools employed, reported effects on student learning outcomes, and factors influencing integration. The review followed the PRISMA 2020 framework. Searches of PubMed, Web of Science, and Scopus covered peer-reviewed English-language studies published from January 2021 to January 27, 2026. The search identified 20,671 records; 2,615 remained after database filtering, 188 were sought for retrieval, and 170 full texts were assessed. Joanna Briggs Institute appraisal classified 84 studies as high quality, 74 as medium quality, and 12 as low quality; the low-quality studies were excluded, leaving 158 studies for synthesis. ChatGPT was the most frequently examined tool, appearing in 112 studies (70.9%). Academic writing and essay support was the most common application area (28 studies, 17.7%), followed by personalised tutoring and feedback (22 studies, 13.9%). Seventy-two studies found positive effects on academic performance, while other studies reported non-significant or negative outcomes. Digital and AI literacy findings were predominantly positive, whereas evidence on critical thinking, creativity, and metacognitive skills was mixed. Effective integration was associated with individual readiness, ethical and social considerations, institutional support, pedagogical design, and technical infrastructure. Overall, outcomes depend on purposeful, supported, and critically engaged use. The review highlights the importance of contextualised institutional guidance, professional development, ethical safeguards, and adequate infrastructure for sustainable educational use. These findings suggest that access to GenAI tools alone does not guarantee improved learning; educational value depends on how tools are integrated into pedagogy, assessment, learner support, and critical engagement.

Keywords: Generative Artificial Intelligence, Higher Education, ChatGPT, Learning Outcomes, AI Literacy

INTRODUCTION

Generative Artificial Intelligence (GenAI) has become a significant development in higher education, particularly since the public emergence of large language model applications such as ChatGPT. These systems can generate and transform text, provide explanations, support academic writing, offer feedback, assist with coding, and respond conversationally to learner questions. Their rapid uptake has therefore expanded the range of activities through which artificial intelligence can participate in teaching and learning. At the same time, the educational use of GenAI has raised questions about academic integrity, reliability, originality, bias, privacy, assessment, and students' capacity to remain active and critical learners (Bhullar et al., 2024). The growing literature reflects both opportunities and concerns, but it is dispersed across disciplines, technologies, educational purposes, and research designs.

The emerging evidence suggests that GenAI can function as a form of learning support rather than simply as an automated answer generator. Students may use conversational systems to obtain explanations, receive feedback, brainstorm ideas, revise written work, practise problem solving, or obtain support during independent study. Instructors may also use GenAI for assessment design, instructional materials, feedback, and other teaching tasks (Gupta, 2024). Reviews of the field have reported a broad range of applications and have emphasised the need to understand how these applications interact with pedagogical design and learner characteristics (Yusuf et al., 2024; Mohebi, 2024; Cai et al., 2025). However, the availability of a GenAI tool does not itself establish a learning benefit. The educational value of an AI response depends on whether students evaluate its accuracy, connect it to disciplinary knowledge, and use it in ways that support rather than replace learning processes.

Learning outcomes are particularly important because GenAI can improve task efficiency while producing less certain effects on deeper learning. Evidence in the reviewed literature includes improvements in academic performance, knowledge acquisition, engagement, digital literacy, and self-regulated learning, but it also includes mixed and negative findings. Fan et al. (2025), for example, identified risks of "metacognitive laziness" when learners rely on GenAI without sufficient cognitive engagement. Similarly, Xu et al. (2025) showed that metacognitive support can shape whether GenAI environments strengthen or weaken self-regulated learning. These findings suggest that outcome differences are not adequately explained by the presence of AI alone. They require attention to the nature of the task, the learner's prior competence, the design of the learning activity, and the type of support provided by teachers and institutions.

The literature also highlights the importance of responsible and context-sensitive integration. Ethical issues include privacy, academic integrity, fairness, transparency, trust, and the reliability of generated information (Omeh et al., 2025; Kodra et al., 2025). Institutional policies and professional development can influence how educators perceive and use GenAI, particularly for assessment (Khlaif et al., 2024; Cabellos et al., 2024). Technical conditions matter as well. Access to appropriate devices, connectivity, integrated platforms, and usable tools can shape the feasibility of adoption, while students' digital competence and AI literacy can influence the quality of information they obtain and how critically they use it (Qi et al., 2024). Thus, GenAI integration is better understood as an educational and institutional process than as a technology-adoption decision alone.

Despite rapid growth in research, several gaps remain. First, much of the literature has concentrated on ChatGPT, leaving newer systems such as Gemini, Claude, Copilot, DeepSeek, and other models less examined. Second, quantitative studies dominate the evidence base, while qualitative studies that explore how students and educators experience GenAI in particular contexts are less common (Haroud & Saqri, 2025). Third, the geographical distribution of studies is uneven, with substantial evidence from China, the United States, Australia, Turkey, and the United Kingdom, but comparatively less evidence from low-resource settings. These gaps limit the extent to which findings can be transferred across institutional and socioeconomic contexts.

Relative to existing syntheses of GenAI in education (Yusuf et al., 2024; Mohebi, 2024; Cai et al., 2025), this review makes three distinct contributions. First, it integrates tools and learning purposes, learning outcomes, and integration factors within a single evidence base that passed explicit quality appraisal, whereas most prior reviews address only one of these dimensions. Second, it reports outcome patterns by domain with explicit denominators, distinguishing measured learning gains from self-reported perceptions. Third, it connects outcome patterns to the individual, ethical, institutional, and technical conditions of implementation, thereby clarifying not only what has been studied but under what conditions GenAI use appears educationally productive.

This systematic review responds to these gaps by bringing together evidence on three connected dimensions: the GenAI tools being used in higher education and their educational purposes; their reported effects on student learning outcomes; and the individual, ethical, institutional, pedagogical, and technical factors that influence integration. The review addresses the following research questions:

RQ1: What Generative AI tools are being used in higher education, and for what learning purposes?

RQ2: What effects do Generative AI tools have on student learning outcomes, including academic performance, engagement, and skill development?

RQ3: What factors influence the effective integration of Generative AI tools in higher education?

By synthesising these dimensions within one evidence base, the review provides a structured account of what has been studied, what outcomes have been reported, and what conditions appear important for educationally meaningful use. The synthesis is intended to support researchers, educators, and institutions seeking to interpret the rapidly developing GenAI literature without assuming that technological adoption automatically produces educational improvement. The literature therefore calls for synthesis that connects tool use with outcomes and implementation conditions. A review organised around these three dimensions can distinguish frequently studied applications from less examined ones and can clarify where findings converge or diverge. It can also identify methodological and geographical patterns that affect the strength and transferability of current evidence.

METHODOLOGY

The review followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 framework to identify, screen, assess, and synthesise relevant evidence (Page et al., 2021). The review protocol was organised around the three research questions and focused on empirical research concerning GenAI in higher education. The final database search was conducted on January 27, 2026.

Search strategy and information sources.

Systematic searches were conducted in PubMed, Web of Science, and Scopus. The search strategy combined three concept blocks: GenAI, higher education, and learning outcomes/integration factors. GenAI terms included "generative artificial intelligence," "generative AI," "GenAI," "large language model," "LLM," "foundation model," "ChatGPT," "GPT-3," "GPT-4," "Gemini," "Claude," "Copilot," "chatbot," "AI tutor," "AI-generated content," and related terms. Higher education terms included "higher education," "tertiary education," "university," "college," "undergraduate," "postgraduate," and "graduate education." Outcome and integration terms included academic performance, learning outcomes, student achievement, skill development, critical thinking, problem solving, creativity, engagement, motivation, academic integrity, plagiarism, assessment, pedagogical strategy, curriculum integration, ethical implications, bias, fairness, digital divide, policy, and guidelines. The three concept blocks were combined using AND, while terms within each block were combined using OR. Citation searching was also used where additional relevant studies were identified.

Table 1a: Database-specific search strategies, search fields, and filters

Database	Search fields	Search string (combined as a AND b AND c)	Filters applied	Records (n)
PubMed	Title/Abstract	("generative artificial intelligence"[Title/Abstract] OR "generative AI"[Title/Abstract] OR GenAI[Title/Abstract] OR "large language model"[Title/Abstract] OR LLM[Title/Abstract] OR ChatGPT[Title/Abstract] OR "GPT-3"[Title/Abstract] OR "GPT-4"[Title/Abstract] OR Gemini[Title/Abstract] OR Claude[Title/Abstract] OR Copilot[Title/Abstract] OR chatbot[Title/Abstract] OR "AI tutor"[Title/Abstract]) AND ("higher education"[Title/Abstract] OR "tertiary education"[Title/Abstract] OR university*[Title/Abstract] OR college*[Title/Abstract] OR undergraduate*[Title/Abstract] OR postgraduate*[Title/Abstract] OR "graduate education"[Title/Abstract])	English; 1 Jan 2021–27 Jan 2026; peer-reviewed journal articles	428
Web of Science	Topic (TS=)	TS=(("generative artificial intelligence" OR "generative AI" OR GenAI OR "large language model" OR LLM OR ChatGPT OR "GPT-3" OR "GPT-4" OR Gemini OR Claude OR Copilot OR chatbot OR "AI tutor" OR "AI-generated content")	English; 1 Jan 2021–27 Jan 2026; article document type	19,433

		AND (“higher education” OR “tertiary education” OR university* OR college* OR undergraduate* OR postgraduate* OR “graduate education”))		
Scopus	TITLE-ABS-KEY	TITLE-ABS-KEY((“generative artificial intelligence” OR “generative AI” OR GenAI OR “large language model” OR LLM OR ChatGPT OR “GPT-3” OR “GPT-4” OR Gemini OR Claude OR Copilot OR chatbot OR “AI tutor” OR “AI-generated content”) AND (“higher education” OR “tertiary education” OR university* OR college* OR undergraduate* OR postgraduate* OR “graduate education”))	English; 2021–2026; source type: journals	810

The reduction from 20,671 identified records to 2,615 records screened reflects database-level filters applied at the time of searching: English language, publication between January 2021 and January 27, 2026, and peer-reviewed article document types. These filters removed 18,056 records before screening. Because filtering was applied within each database, no additional duplicate-removal step was required before screening

Eligibility and screening.

We included peer-reviewed full research articles in English, published between January 2021 and January 27, 2026, indexed in at least one of the three databases, focused on higher education, and directly examined GenAI, its educational use, student learning, academic performance, or factors influencing integration. We excluded review articles, books, book chapters, editorials, meeting abstracts, data papers, dissertations, preprints, conference abstracts, studies without retrievable full text, and studies outside the defined context. The initial search produced 20,671 records. After database filters, 2,615 records remained for title and abstract screening. Three researchers independently assessed the records and reported 95% agreement. A total of 188 reports were sought for retrieval; 18 could not be retrieved, leaving 170 reports for full-text eligibility assessment and quality appraisal.

Under these criteria, systematic reviews and conceptual or theoretical papers were not eligible for outcome synthesis. Six systematic reviews and five conceptual or theoretical papers identified during screening were retained in the corpus solely as background and contextual references; they are identified in Supplementary Material S1 and were excluded from all outcome counts and from the thematic synthesis (see Results and the note to Table 5).

Quality appraisal and synthesis.

Methodological quality was assessed with the Joanna Briggs Institute critical appraisal frameworks, selecting the tool appropriate to each study design (Aromataris & Munn, 2020). Quality was operationalised as high when at least 80% of applicable criteria were met, medium at 60–79%, and low below 60%. The appraisal identified 84 high-quality, 74 medium-quality, and 12 low-quality studies. The 12 low-quality studies were excluded, producing a final analytical corpus of 158 studies. Two researchers independently coded the extracted evidence, harmonised their coding into themes, and used a third researcher to verify the themes. The synthesis combined thematic analysis following Braun and Clarke’s (2006) approach with narrative synthesis to identify patterns across studies and relate them to the three research questions (Popay et al., 2006).

Search Concepts and Boolean Operators

Methodological quality was assessed using the Joanna Briggs Institute (JBI) critical appraisal checklists, with the instrument selected to match each study's design (Aromataris & Munn, 2020): cross-sectional surveys were appraised with the JBI checklist for analytical cross-sectional studies; randomised trials with the JBI checklist for randomised controlled trials; non-randomised experiments with the JBI checklist for quasi-experimental studies; longitudinal designs with the JBI checklist for cohort studies; and qualitative studies with the JBI checklist for qualitative research. Where a study design did not map directly onto one of these instruments, the

most appropriate JBI checklist was applied, and the choice was recorded. Complete study-level JBI assessments, the checklist applied, the items met, and the resulting rating for each of the 170 full-text reports assessed are reported in Supplementary Material S1, so that every quality judgement underlying the synthesis is traceable to item-level evidence.

Quality was operationalised as high when at least 80% of the applicable checklist items were met, medium when 60–79% were met, and low when fewer than 60% were met; items judged not applicable to a given study were removed from the denominator. These thresholds were fixed a priori in the review protocol before screening or appraisal began, so that no quality decision could be influenced by the results of the studies under review. They are also consistent with the cut-offs used in recent systematic reviews of generative AI in education, in which studies scoring below 60% were likewise excluded from synthesis. The 12 studies rated low quality were therefore excluded from the analytical corpus (n = 158) to reduce the risk that weak designs would distort the outcome patterns reported below.

Table 1: Search concepts, terms, and Boolean structure

Concept	Principal search terms
Generative AI (GenAI)	"Generative artificial intelligence" OR "generative AI" OR "GenAI" OR "large language model" OR "LLM" OR "foundation model" OR "ChatGPT" OR "GPT-3" OR "GPT-4" OR "Gemini" OR "Claude" OR "Copilot" OR "chatbot" OR "AI tutor" OR "AI-generated content"
Higher Education	"Higher education" OR "tertiary education" OR "university" OR "college" OR "undergraduate" OR "postgraduate" OR "graduate education"
Learning Outcomes and Integration	"Learning outcome" OR "academic performance" OR "student achievement" OR "grade" OR "skill development" OR "critical thinking" OR "problem solving" OR "creativity" OR "student engagement" OR "motivation" OR "academic integrity" OR "plagiarism" OR "assessment" OR "pedagogical strategy" OR "ethical implication" OR "bias" OR "fairness" OR "policy" OR "guideline"

Note: The three concept blocks were combined as (a) AND (b) AND (c), with OR used between terms within each block.

Eligibility Criteria

Table 2: Inclusion and exclusion criteria

Criterion	Inclusion	Exclusion
Language	Research articles published in English	Articles not published in English
Publication type	Peer-reviewed full research articles	Reviews, books, chapters, editorials, abstracts, data papers, dissertations, preprints, conference abstracts
Quality	High- or medium-quality appraisal score	Low-quality appraisal score
Full text	Full text retrievable	Full text unavailable
Database	Indexed by PubMed, Web of Science, or Scopus	Not indexed by the selected databases
Publication date	January 2021 to January 27, 2026	Before January 2021 or after January 27, 2026
Setting	Higher education	Other educational settings
Scope	GenAI, higher education, learning, outcomes, or integration factors	Outside the review scope

Quality Appraisal and Synthesis

Table 3: Study selection and methodological quality appraisal

Stage	Number of records/studies
Initial database records	20,671
After database filtering	2,615
Reports sought for retrieval	188
Full texts unavailable	18
Full texts assessed	170
High-quality studies	84
Medium-quality studies	74
Low-quality studies excluded	12
Final studies included	158

RESULTS AND DISCUSSION

The final synthesis included 158 studies. Only 17 included studies were published between 2021 and 2023, compared with 80 in 2024 and 61 in 2025. No eligible study in the final corpus was dated up to January 27, 2026. The geographical analysis indicated that China and the United States contributed the largest numbers of studies, followed by Australia, Turkey, and the United Kingdom, with research distributed across more than 30 countries.

RQ1: GenAI tools and learning purposes. ChatGPT was the dominant tool in the literature, occurring in 112 of the 158 studies (70.9%). Gemini/Bard appeared in eight studies (5.1%), Microsoft Copilot in six (3.8%), Claude in four (2.5%), Llama in three (1.9%), DeepSeek in two (1.3%), and Perplexity AI in two (1.3%). Eighteen studies (11.4%) referred to GenAI without identifying a specific platform. The predominance of ChatGPT is consistent with earlier reviews that identify it as the most visible GenAI application in education (Bhullar et al., 2024; Mohebi, 2024). The concentration also creates a research gap because evidence about newer tools is still comparatively small.

Academic writing and essay support was the most frequently reported application area, accounting for 28 studies (17.7%). Personalised tutoring and feedback followed with 22 studies (13.9%), while medical and health professions education accounted for 18 studies (11.4%). Problem solving and clinical reasoning appeared in 14 studies (8.9%), STEM and mathematics in 10 (6.3%), language learning in eight (5.1%), assessment design and feedback in seven (4.4%), programming and computer science in six (3.8%), and research support in five (3.2%). These findings show that GenAI research has concentrated strongly on text-based and knowledge-intensive activities. The smaller evidence base in programming, language learning, and STEM also points to the need for further investigation of discipline-specific uses and limitations (Suh et al., 2025).

RQ2: Learning outcomes. The review identified 72 studies reporting positive effects on academic performance, including grade improvement, knowledge retention, or task efficiency. This pattern is consistent with research suggesting that GenAI can provide scaffolding and timely support when its use is aligned with learning tasks (Pallant et al., 2025; Xia et al., 2025). However, the evidence was not uniformly positive. Nine studies reported no significant difference, and seven reported negative effects. Fan et al. (2025) reported that students could obtain higher short-term essay performance without corresponding gains in knowledge transfer or intrinsic motivation, while Azeem and Abbas (2025) reported relationships between academic GenAI use, self-efficacy, and learned helplessness. The findings therefore indicate that task performance and learning are not necessarily equivalent outcomes.

In total, 88 of the 158 studies assessed academic performance as an outcome (72 positive, nine reporting no significant difference, and seven negative); the remaining 70 studies did not assess academic performance and are therefore not represented in these counts.

Engagement and motivation showed a similar pattern. Thirty-four studies reported positive effects, four reported mixed effects, and one reported a negative effect. AI-generated virtual instructors, for example, were associated with increased affective engagement in one study (Chen et al., 2025). In contrast, Durgungoz and Durgungoz (2025) observed that an initial increase in engagement with AI-generated gamified quizzes declined over time. Skill-development evidence was also heterogeneous. Digital and AI literacy showed the most consistently positive pattern, with 10 positive studies, one mixed/no-effect study, and no negative study in the extracted table. Self-regulated learning included 12 positive, three mixed/no-effect, and four negative studies. Critical thinking showed eight positives, six mixed/no-effect, and five negative studies; creativity showed four positives, five mixed/no-effect, and three negative studies. Metacognitive awareness was the most concerning domain, with three positive and six negative studies. These findings are consistent with the argument that explicit metacognitive support can reduce risks of dependency and strengthen self-regulated learning (Xu et al., 2025; Shomotova et al., 2026).

In total, 39 of the 158 studies assessed engagement or motivation (34 positive, four mixed, and one negative); engagement or motivation was not assessed in the remaining studies.

RQ3: Factors influencing integration. Four broad categories emerged from the synthesis. Individual and psychological factors were identified in 48 studies and included attitudes, digital self-efficacy, anxiety, prior experience, perceived responsiveness, and enjoyment. Faculty anxiety and concerns about misuse were associated with adoption intentions in one study (Verano-Tacoronte et al., 2025), while usability and enjoyment were linked to student motivation and satisfaction in another (Kim et al., 2025). Ethical and social factors appeared in 42 studies and included privacy, academic integrity, fairness, transparency, trust, and perceived limitations. Institutional policies and support were identified in 35 studies, with evidence emphasising clear guidelines, assessment redesign, administrative support, and training (Khlaif et al., 2024; Cabellos et al., 2024; Xia et al., 2024. Technical infrastructure and design appeared in 28 studies, including accessibility, usability, task-technology fit, and platform integration. Al-Emran et al. (2025) showed that task and technology characteristics can shape task-technology fit, while Qi et al. (2024) highlighted the importance of critical thinking, self-directed learning, and AI literacy in determining the quality of information students obtain from GenAI.

Methodologically, quantitative studies dominated the corpus (82; 52%), followed by mixed methods (35; 22%) and qualitative studies (14; 9%). Of the 147 empirical studies, 16 used other or unspecified approaches (10%). The six systematic reviews (4%) and five conceptual or theoretical papers (3%) were retained solely as background and contextual references, consistent with the eligibility criteria; they are identified in Supplementary Material S1 and were excluded from all outcome counts and the thematic synthesis. The dominance of quantitative designs provides useful evidence about measurable outcomes but leaves less detailed evidence about how educators and students experience GenAI in particular institutional and cultural contexts. Geographical concentration similarly affects transferability. The strong representation of China and the United States contrasts with the comparatively limited evidence from low-income settings, where infrastructure, connectivity, institutional resources, and AI literacy may differ substantially.

Across the five skill domains, 70 domain assessments were reported: digital/AI literacy (11 studies), self-regulated learning (19), critical thinking (19), creativity (12), and metacognitive awareness (9). Some studies assessed more than one domain; a domain not assessed by a study is not represented in that study's counts.

The outcome evidence was also differentiated by evidential strength. Where findings relied on self-reported perceptions (for example, perceived learning, confidence, or satisfaction) rather than measured learning gains (for example, test performance or graded work), they are interpreted as perception evidence rather than demonstrated learning effects; the evidential basis of each outcome is documented in Supplementary Material S1. Interpretation of the outcome patterns also took account of sample size, study design, the presence or absence of a comparator condition, and follow-up duration.

In addition, the association between digital competence and AI literacy reported by Shomotova et al. (2026) is correlational and should not be interpreted as evidence that GenAI use itself improves these competencies; rather,

pre-existing digital competence and critical thinking may shape both AI literacy and the quality of the information students obtain from GenAI (Qi et al., 2024).

The findings summarised above constitute the verified results of the synthesis. The interpretive statements and recommendations that follow are broader inferences drawn from these findings and are deliberately separated here from the verified results.

Taken together, the findings indicate that GenAI integration is multidimensional. The evidence does not support treating a particular tool as inherently educationally beneficial or harmful. Instead, outcomes vary with the purpose of use, learner characteristics, pedagogical design, institutional support, ethical safeguards, and technical conditions. The review is also subject to limitations. It included only English-language peer-reviewed studies indexed in three databases, excluded grey literature, covered a rapidly changing period, and combined heterogeneous study designs and outcome measures that did not permit a formal meta-analysis. Quality appraisal involved researcher judgement despite the use of explicit JBI criteria. These limitations should be considered when interpreting the scope and transferability of the synthesis.

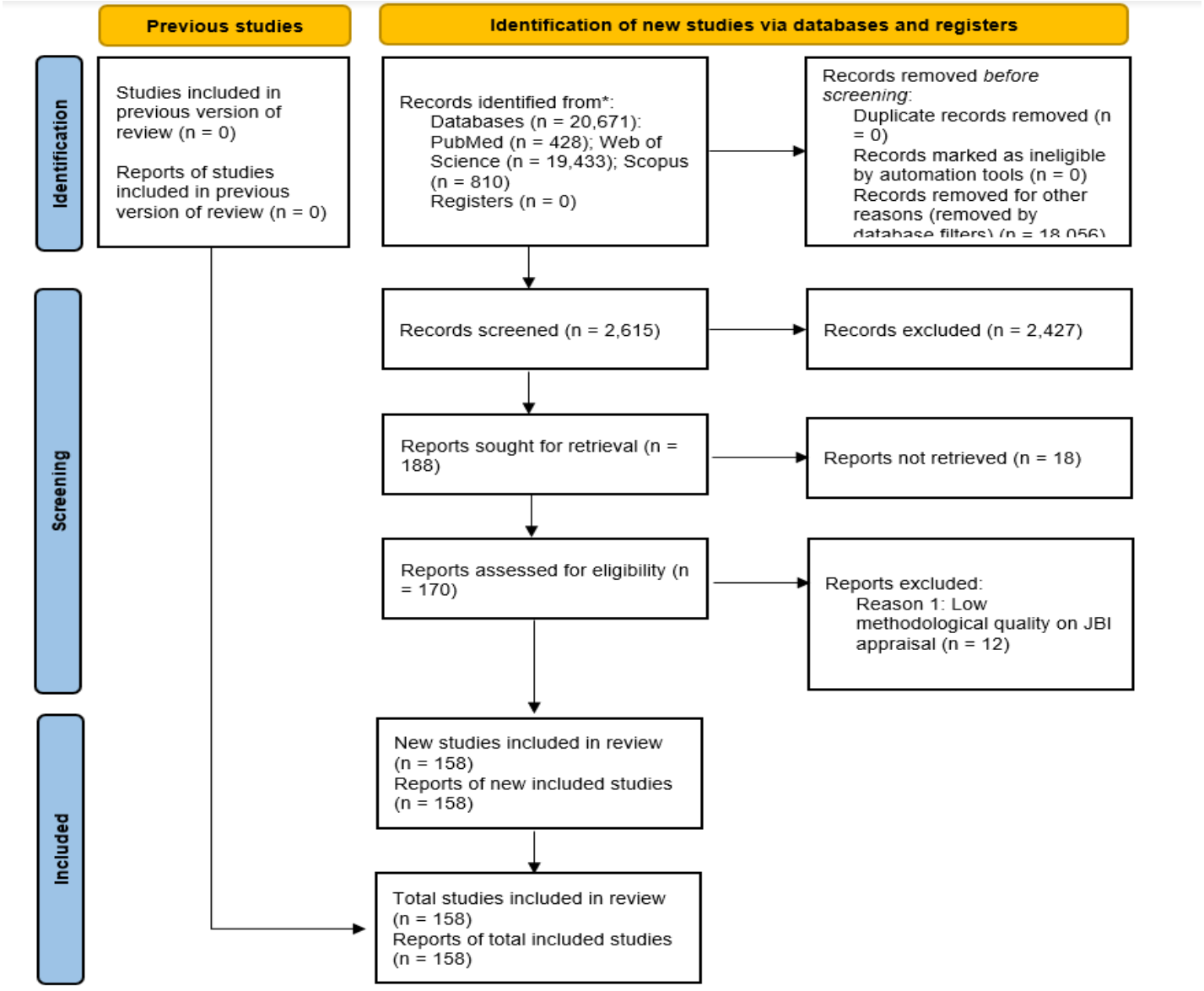


Figure 1:Prisma 2020 flowchart for study identification, screening, eligibility, and inclusion

Table 4: Annual distribution of included publications, 2021–2026

Year	Number of publications	Cumulative total
2021	3	3

2022	10	13
2023	4	17
2024	80	97
2025	61	158
2026 (to Jan 27)	0	158
Total	158	—

Note: Publication output peaked at 80 studies in 2024 and then declined to 61 studies in 2025; the 2025 output nevertheless remained substantially higher than the 17 studies published across 2021–2023. (Corrected: the previous version of this note did not acknowledge the decline from 80 to 61 publications.)

Table 5: Distribution of methodological approaches

Methodological approach	**n**	**Percentage**
Quantitative	82	52%
Mixed methods	35	22%
Qualitative	14	9%
Systematic reviews	6	4%
Conceptual/theoretical	5	3%
Not specified/other	16	10%
Total	158	100%

Note: The corpus comprised 147 empirical studies contributing to the synthesis (quantitative, mixed methods, qualitative, and other/unspecified approaches) and 11 background references (six systematic reviews and five conceptual/theoretical papers) that were excluded from outcome counts and thematic synthesis.

Table 5a: Alignment of research questions, themes, and codes used in the thematic synthesis

Research question	**Theme**	**Illustrative codes**
RQ1: Tools and learning purposes	GenAI tools in use	ChatGPT; Gemini/Bard; Microsoft Copilot; Claude; Llama; DeepSeek; Perplexity AI; unspecified GenAI tools
RQ1: Tools and learning purposes	Learning purposes / application areas	Academic writing and essay support; personalised tutoring and feedback; problem-solving and clinical reasoning; language learning; programming/computer science; assessment design and feedback; research support
RQ2: Learning outcomes	Academic performance	Grade improvement; knowledge retention; task efficiency; no significant difference; negative effects
RQ2: Learning outcomes	Engagement and motivation	Affective, behavioural, and cognitive engagement; sustained motivation; engagement decline over time
RQ2: Learning outcomes	Skill development	Digital/AI literacy; self-regulated learning; critical thinking; creativity; metacognitive awareness
RQ3: Integration factors	Individual and psychological factors	Attitudes; digital self-efficacy; anxiety; prior experience; perceived responsiveness; enjoyment
RQ3: Integration factors	Ethical and social factors	Privacy; academic integrity; fairness; transparency; trust; perceived limitations
RQ3: Integration factors	Institutional and support factors	Guidelines; assessment redesign; administration; training
RQ3: Integration factors	Technical infrastructure and design	Accessibility; usability; task-technology fit; platform integration

Table 6: Generative AI tools represented in the included studies

GenAI tool	Occurrences	Percentage of studies
ChatGPT (GPT-3.5/GPT-4/GPT-4o)	112	70.9%
Unspecified GenAI tools	18	11.4%
Gemini/Bard	8	5.1%
Microsoft Copilot	6	3.8%
Claude	4	2.5%
Llama	3	1.9%
DeepSeek	2	1.3%
Perplexity AI	2	1.3%
Other (single occurrence each)	3	1.9%
Total	158	100%

Table 7: Application areas of Generative AI in higher education research

Application area	Studies	Percentage
Academic writing and essay support	28	17.7%
Personalised tutoring and feedback	22	13.9%
Medical and health professions education	18	11.4%
Problem-solving and clinical reasoning	14	8.9%
STEM/mathematics education	10	6.3%
Language learning	8	5.1%
Assessment design and feedback	7	4.4%
Programming/computer science education	6	3.8%
Research support	5	3.2%
Other/not specified	40	25.3%
Total	158	100%

Note: Application areas were coded by primary educational purpose (for example, writing support, tutoring, or assessment design). Categories are not mutually exclusive; where a study served more than one purpose, the primary purpose reported by the study authors was coded. Disciplinary settings (for example, medical and health professions education and STEM/mathematics education) are listed alongside purposes because a substantial share of GenAI research is organised around professional disciplines.

Table 8: Reported effects of Generative AI on student skill development

Skill domain	Positive	Mixed/no effect	Negative
Digital/AI literacy	10	1	0
Self-regulated learning	12	3	4
Critical thinking	8	6	5
Creativity	4	5	3
Metacognitive awareness	3	0	6

Note: Denominators differ across domains because not all studies assessed every skill; the number of studies contributing to each domain is reported in the text.

Table 9: Main factors influencing effective GenAI integration

Factor category	Studies identifying the category	Examples of factors
Individual and psychological	48	Attitudes, self-efficacy, anxiety, prior experience, responsiveness, enjoyment
Ethical and social	42	Privacy, academic integrity, fairness, transparency, trust

Institutional and support	35	Guidelines, assessment redesign, administration, training
Technical infrastructure and design	28	Accessibility, usability, task-technology fit, platform integration

Note: Factor categories are not mutually exclusive; a single study could contribute to more than one category, so the category counts do not sum to 158.

CONCLUSION

This systematic review synthesised 158 peer-reviewed studies on Generative Artificial Intelligence in higher education published between January 2021 and January 27, 2026. GenAI was examined mainly through ChatGPT, academic writing, personalised tutoring, feedback, and other text-based learning activities. ChatGPT appeared in 112 studies (70.9%), while academic writing and essay support was the most frequently reported application area. The literature has therefore concentrated on a narrow set of tools and uses even as the GenAI landscape expands.

The evidence concerning learning outcomes is encouraging in some areas but heterogeneous. Seventy-two studies reported positive effects on academic performance, and digital and AI literacy showed a predominantly positive pattern. Engagement and motivation were also frequently positive, although some evidence indicated that initial engagement may decline over time. Critical thinking, creativity, self-regulated learning, and particularly metacognitive awareness produced more mixed findings. Several studies identified risks associated with uncritical reliance, including reduced knowledge transfer, dependency, and weaker metacognitive engagement. Thus, efficiency in completing an academic task should not automatically be treated as equivalent to durable learning. GenAI may support learning when it is embedded in activities that require students to question, verify, explain, compare, revise, and reflect on AI-generated content.

Effective integration also depends on conditions surrounding the technology. Individual readiness, attitudes, confidence, prior experience, and anxiety influence use. Ethical and social concerns concerning privacy, academic integrity, fairness, transparency, and trust shape adoption. Institutional policies, professional development, assessment design, and administrative support provide an organisational framework, while infrastructure, accessibility, usability, and task-technology fit determine practical feasibility. These factors are interconnected. Training, for example, can strengthen users' ability to evaluate AI outputs, while assessment policies can influence whether GenAI becomes pedagogically productive or simply a shortcut.

The recommendations that follow are interpretive inferences derived from the synthesised evidence rather than verified findings. For higher education institutions, the evidence supports a contextual approach. Institutions should provide clear guidance on acceptable use, academic integrity, transparency, and verification of AI-generated information. Professional development should include AI literacy, critical evaluation, ethical decision making, and pedagogical design. Assessment practices can place greater emphasis on learning processes, explanation, reflection, and application rather than final written products alone. Educators can provide metacognitive scaffolding so that students use GenAI as a resource for thinking rather than a substitute for thinking. Technology developers should also consider source transparency, reliability, citation support, and features that encourage self-regulated learning (Bawn et al., 2026; Kodra et al., 2025).

Future research should include longitudinal studies, qualitative and mixed-methods research, comparisons of emerging GenAI tools, and more evidence from low-resource contexts. Discipline-specific research and intervention studies are also needed to determine which pedagogical approaches work, for whom, and under what conditions. Research on sustained teaching and learning innovation can also help institutions understand how GenAI-related practices become embedded over time (Bearman et al., 2024). Overall, the evidence presents GenAI as a flexible educational technology whose effects depend on purpose, pedagogy, learner characteristics, institutional support, ethical safeguards, and technical conditions. Its integration should therefore be guided by explicit learning objectives, critical engagement, appropriate assessment, learner autonomy, and responsible use rather than by technology access alone.

REFERENCES

1. Al-Emran, M., Al-Qaysi, N., Al-Sharafi, M. A., Khoshkam, M., Foroughi, B., & Ghobakhloo, M. (2025). Role of perceived threats and knowledge management in shaping generative AI use in education and its impact on social sustainability. The International Journal of Management Education, 23(1), 101105. https://doi.org/10.1016/j.ijme.2024.101105
2. Aromataris, E., & Munn, Z. (Eds.). (2020). JBI manual for evidence synthesis. JBI. https://doi.org/10.46658/JBIMES-20-01
3. Azeem, S., & Abbas, M. (2025). Personality correlates of academic use of generative artificial intelligence and its outcomes: Does fairness matter? Education and Information Technologies, 30, 18131–18155. https://doi.org/10.1007/s10639-025-13489-6
4. Bawn, M., Francis, N., Alvey, E., Hassall, C., Pires-daSilva, A., Barra, P., Hough, D., Campbell, H., Hardy, M., & Canet-Perez, J. (2026). Perspectives from a workshop: Intelligent assessment in the age of artificial intelligence. Advances in Physiology Education, 50(1), 73–82. https://doi.org/10.1152/advan.00246.2024
5. Bearman, M., Chandir, H., Mahoney, P., & Partridge, H. (2024). Sustaining teaching and learning innovations: A scoping review. Higher Education Research & Development, 43(7), 1495–1510. https://doi.org/10.1080/07294360.2024.2364096
6. Bhullar, P. S., Joshi, M., & Chugh, R. (2024). ChatGPT in higher education: A synthesis of the literature and a future research agenda. Education and Information Technologies, 29, 21501–21522. https://doi.org/10.1007/s10639-024-12723-x
7. Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative Research in Psychology, 3(2), 77–101. https://doi.org/10.1191/1478088706qp063oa
8. Cabellos, B., de Aldama, C., & Pozo, J.-I. (2024). University teachers' beliefs about the use of generative artificial intelligence for teaching and learning. Frontiers in Psychology, 15, 1468900. https://doi.org/10.3389/fpsyg.2024.1468900
9. Cai, C., Zhu, G., & Ma, M. (2025). A systemic review of AI for interdisciplinary learning: Application contexts, roles, and influences. Education and Information Technologies, 30(7), 9641–9687. https://doi.org/10.1007/s10639-024-13193-x
10. Chen, J., Mokmin, N. A. M., Shen, Q., & Su, H. (2025). Leveraging AI in design education: Exploring virtual instructors and conversational techniques in flipped classroom models. Education and Information Technologies, 30, 16441–16461. https://doi.org/10.1007/s10639-025-13458-z
11. Durgungoz, A., & Durgungoz, F. C. (2025). Exploring effortless AI-generated gamified quizzes in an online special education module: Evaluating question quality, student engagement, and its potential to identify at-risk students. Education and Information Technologies, 30, 25335–25357. https://doi.org/10.1007/s10639-025-13765-5
12. Fan, Y., Tang, L., Le, H., Shen, K., Tan, S., Zhao, Y., Shen, Y., Li, X., & Gašević, D. (2025). Beware of metacognitive laziness: Effects of generative artificial intelligence on learning motivation, processes, and performance. British Journal of Educational Technology, 56(2), 489–530. https://doi.org/10.1111/bjet.13544
13. Gupta, A. (2024). Impact of generative AI in transforming higher education pedagogy. In R. Bansal, A. Ngah, A. Chakir, & N. Pruthi (Eds.), Leveraging ChatGPT and artificial intelligence for effective customer engagement (pp. 285–300). IGI Global Scientific Publishing. https://doi.org/10.4018/979-8-3693-0815-8.ch017
14. Haroud, S., & Saqri, N. (2025). Generative AI in higher education: Teachers' and students' perspectives on support, replacement, and digital literacy. Education Sciences, 15(4), 396. https://doi.org/10.3390/educsci15040396
15. Khlaif, Z. N., Ayyoub, A., Hamamra, B., Bensalem, E., Mitwally, M. A. A., Ayyoub, A., Hattab, M. K., & Shadid, F. (2024). University teachers' views on the adoption and integration of generative AI tools for student assessment in higher education. Education Sciences, 14(10), 1090. https://doi.org/10.3390/educsci14101090
16. Kim, M., Kim, J., Knotts, T. L., & Albers, N. D. (2025). AI for academic success: Investigating the role of usability, enjoyment, and responsiveness in ChatGPT adoption. Education and Information Technologies, 30(10), 14393–14414. https://doi.org/10.1007/s10639-025-13398-8

17. Kodra, J. D., Saroyan, A., Darby, F., Surucu, S., Fong, S., Gillinov, S., Girardi, K., Vasudevan, R., Ansah-Twum, J. K., Atadja, L., Moran, J., & Jimenez, A. E. (2025). ChatGPT-generated responses across orthopaedic sports medicine surgery vary in accuracy, quality, and readability: A systematic review. Arthroscopy, Sports Medicine, and Rehabilitation, 7(4), 101210. https://doi.org/10.1016/j.asmr.2025.101210
18. Mohebi, L. (2024). Empowering learners with ChatGPT: Insights from a systematic literature exploration. Discover Education, 3, 36. https://doi.org/10.1007/s44217-024-00120-y
19. Omeh, C. B., Olelewe, C. J., & Hu, X. (2025). Application of artificial intelligence (AI) technology in TVET education: Ethical issues and policy implementation. Education and Information Technologies, 30(5), 5989–6018. https://doi.org/10.1007/s10639-024-13018-x
20. Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., McGuinness, L. A., Stewart, L. A., Thomas, J., Tricco, A. C., Welch, V. A., 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
21. Pallant, J. L., Blijlevens, J., Campbell, A., & Jopp, R. (2025). Mastering knowledge: The impact of generative AI on student learning outcomes. Studies in Higher Education. Advance online publication. https://doi.org/10.1080/03075079.2025.2487570
22. Popay, J., Roberts, H., Sowden, A., Petticrew, M., Arai, L., Rodgers, M., Britten, N., Roen, K., & Duffy, S. (2006). Guidance on the conduct of narrative synthesis in systematic reviews: A product from the ESRC Methods Programme. University of Lancaster.
23. Qi, C., Tang, Y., & Lei, Y. (2024). Does feedback from ChatGPT help? Investigating the effect of feedback from both teacher and ChatGPT on students' learning outcomes. AMCIS 2024 Proceedings, Article
24. Association for Information Systems. https://aisel.aisnet.org/amcis2024/is_education/is_education/29
25. Shomotova, A., ElSayary, A., & Husain, S. (2026). What shapes students' AI literacy? Investigating digital competence, student background, and GenAI use in higher education. Education and Information Technologies, 31, 387–418. https://doi.org/10.1007/s10639-025-13832-x
26. Suh, J., Lee, K., & Lee, J. (2025). Programming education with ChatGPT: Outcomes for beginners and intermediate students. Education and Information Technologies, 30(14), 19511–19536. https://doi.org/10.1007/s10639-025-13542-4
27. Verano-Tacoronte, D., Bolívar-Cruz, A., & Sosa-Cabrera, S. (2025). Are university teachers ready for generative artificial intelligence? Unpacking faculty anxiety in the ChatGPT era. Education and Information Technologies, 30, 20495–20522. https://doi.org/10.1007/s10639-025-13585-7
28. Xia, Q., Weng, X., Ouyang, F., Lin, T. J., & Chiu, T. K. F. (2024). A scoping review on how generative artificial intelligence transforms assessment in higher education. International Journal of Educational Technology in Higher Education, 21, 40. https://doi.org/10.1186/s41239-024-00468-z
29. Xia, Q., Zhang, P., Huang, W., & Chiu, T. K. F. (2025). The impact of generative AI on university students' learning outcomes via Bloom's taxonomy: A meta-analysis and pattern mining approach. Asia Pacific Journal of Education. Advance online publication. https://doi.org/10.1080/02188791.2025.2530503
30. Xu, X., Qiao, L., Cheng, N., Liu, H., & Zhao, W. (2025). Enhancing self-regulated learning and learning experience in generative AI environments: The critical role of metacognitive support. British Journal of Educational Technology, 56(5), 1842–1863. https://doi.org/10.1111/bjet.13599
31. Yusuf, A., Pervin, N., Román-González, M., & Noor, N. M. (2024). Generative AI in education and research: A systematic mapping review. Review of Education, 12, e3489. https://doi.org/10.1002/rev3.3489