International Journal of Scientific Research & Engineering Trends
Volume 12, Issue 2, Mar-Apr-2026, ISSN (Online): 2395-566X

Generative Artificial Intelligence in Education: A Systematic Literature Review

Tushar Chaudhari
KCES' College of Engineering and Management, Jalgaon

Abstract- The public release of ChatGPT in late 2022 marked a turning point in the adoption and academic investigation of Generative Artificial Intelligence (GenAI). This systematic literature review synthesises 39 peer-reviewed studies to evaluate the applications, pedagogical benefits, and governance challenges of GenAI across global educational contexts. The findings identify four primary application domains: personalised intelligent tutoring, automated content creation, multimodal learning materials, and academic research assistance. Synthesis of the evidence reveals substantial improvements in student performance and affective-motivational states, particularly through adaptive scaffolding and real-time feedback. However, these benefits are countered by significant risks involving academic integrity, "hallucinations," and the potential for cognitive over-reliance. Parallel to these pedagogical concerns, the evidence base remains heavily concentrated in higher education and high-income regions, leaving critical gaps in K-12 settings and the Global South. This review concludes that while GenAI offers transformative potential for personalised learning, its sustainable integration requires robust institutional policy and longitudinal research into long-term cognitive outcomes.

Keywords – Generative Artificial Intelligence; Large Language Models; ChatGPT; Personalised Learning; Academic Integrity; K-12 Education; Higher Education.

I. INTRODUCTION

Computer The emergence of Generative Artificial Intelligence (GenAI) has reshaped education globally. Since ChatGPT launched in November 2022, educators, institutions and students have had to renegotiate the role of machine-generated support in learning. Within two months, ChatGPT reached 100 million monthly users [1][2]. Academic publications on GenAI in education grew from 6 in 2022 to 1,951 in 2024 [3].

GenAI covers a range of architectures: Large Language Models (LLMs) like GPT-4, Claude and Gemini; diffusion-based image generators like DALL-E; and multimodal systems combining text, speech and image [6][1]. Unlike rule-based predecessors, these systems produce contextually coherent outputs from open-ended prompts, enabling deployment across tutoring, assessment and content creation [7][8].

Documented benefits include improved academic performance [2], motivation [9][10] and personalised learning at scale [7][11]. Well-evidenced challenges include academic integrity concerns [12][13], hallucination [4][14], algorithmic bias [14][11] and cognitive over-reliance [15][16]. The tension between GenAI's transformative potential and its risks has driven rapid growth in policy, governance and pedagogical design research, making it one of the most dynamic and consequential fields in contemporary education scholarship.

The literature remains fragmented. Higher education dominates [17][5][2], K-12 is underrepresented [18][19] and studies from the Global South are scarce [11][20]. This review synthesises 39 peer-reviewed studies through a PRISMA-aligned process to provide a rigorous account of the current state of the field.

II. RESEARCH OBJECTIVES

This review is guided by four research objectives:

1. To identify major GenAI applications in education, including tutoring, content creation, multimedia and research assistance.
2. To analyse how tools such as ChatGPT, Khanmigo and Diffit are integrated into teaching practice across educational levels and disciplines.
3. To examine empirical evidence on benefits and limitations including effects on academic performance, motivation, accessibility and equity.
4. To identify research gaps including K-12 settings, audio-based GenAI, longitudinal outcomes and teacher training frameworks.

III. RESEARCH METHODOLOGY (SYSTEMATIC LITERATURE REVIEW)

Literature Search Strategy and Keywords

This review follows PRISMA principles. Searches were conducted across Scopus, Web of Science, IEEE Xplore, ScienceDirect, PubMed and the ACM Digital Library, supplemented by Google Scholar for grey literature and preprints. Keyword clusters included "Generative AI in education," "Large language models pedagogy" and "ChatGPT higher education."

Eligibility Criteria

Papers were included if published between 2022 and early 2026, focused on generative AI in education, written in English and peer-reviewed. Studies were excluded if they covered non-generative AI, lacked empirical analysis or were duplicates.

Paper Selection Process

Initial searches returned 357 records. After de-duplication, 189 unique records were screened by title and abstract. Of these, 78 passed to full-text review. Final screening yielded 39 primary studies.

Each paper was assessed for clarity of research design, transparency of data collection, relevance to the research objectives and explicitness of limitations. The corpus spans experimental studies, systematic reviews, meta-analyses, bibliometric analyses and longitudinal surveys.

IV. RESULTS AND FINDINGS

Overview of Generative AI in Education

- Defining Generative AI

GenAI systems produce new content by learning patterns from large training datasets. Unlike discriminative models, they generate novel outputs across text, image, audio and video from open-ended prompts [1][8]. Four main architectures underpin educational applications.

Purwono et al. [7] position LLMs such as GPT-4, Claude and Gemini as the foundational models for educational deployment. Bernstein et al. [15] categorise tools into five types: chat models, code assistants, search-integrated models, image generators and custom pedagogical tools.

Evolution of GenAI in Education: The 2022–2026 Period

The ChatGPT launch in November 2022 was the decisive turning point [6][19]. Before 2022, educational AI was dominated by rule-based tutoring systems and early chatbots [22][8]. Publications on GenAI in education grew from 6 in 2022 to 1,951 in 2024 [3], a pace faster than comparable growth in AI for medical diagnostics [4].

Within computing education, 2024 produced more GenAI papers than the two preceding years combined, and harm-related papers grew disproportionately [15]. Ogunleye et al. [5] document growth from 38 papers in 2022 to 273 in 2023. The US (935 articles) leads output, followed by China (508) and Hong Kong (147) [3]. Global South contributions remain significantly underrepresented [20][11].

Synthesis of Reviewed Studies

Table 1 summarises the 39 primary studies reviewed, covering research focus, methodology, key findings and limitations.

Table 1: Synthesis of Reviewed Studies on Generative AI in Education

Author & Year	Research Focus	Methodology	Key Findings	Limitations
Raman et al. [23] (2026)	Theoretical review of GenAI adoption; ADO framework	PRISMA SLR; Scopus; n=130 final studies	UTAUT/TAM dominant frameworks; GenAI enhances personalisation, feedback, SRL; over-reliance and ethical risks documented	Single database; English-language only; no empirical data collection
Dhakal [24] (2025)	Bibliometric scoping of GenAI-Pedagogy nexus in HE	PRISMA-ScR bibliometric; Scopus; n=310 outputs (2023-2025)	40 distinct pedagogical approaches; AI literacy most prominent (n=21); urgent need for policy, curriculum redesign and professional development	Generic search terms; LLM-specific terminology excluded; grey literature excluded
Francis, Jones & Smith [25] (2025)	GenAI impact on assessment and academic integrity in HE	Narrative literature review; secondary analysis of institutional policy surveys	<50% of top universities had public GenAI guidelines; >94% educators report no clear institutional policy; automated grading comparable to human marking	No systematic database search; primarily UK/Anglophone context; no longitudinal data
Ogunleye et al. [26] (2024)	GenAI assessment capability across STEM disciplines	Experimental; ChatGPT-4 and Bard tested on master's-level	GenAI showed subject knowledge and analytical skills; construction management exposed limitations; minimal hallucination	Limited to 3 disciplines at one institution; model capabilities since evolved

		assessments; criterion-referenced marking		
Ilieva et al. [27] (2025)	Framework for GenAI-supported assessment in HE	Conceptual framework; comparative case study; 15 students' final exams graded by human and ChatGPT 4.5	ChatGPT scores aligned in 12 of 15 cases; MAE 0.53-1.87; hybrid human-AI model recommended	Single institution; static dataset; no longitudinal analysis
Benedek & Sziklai [28] (2025)	Impact of uncontrolled AI use on genuine knowledge acquisition	RCT; ~100 students; offline true-or-false paper test + CopyLeaks detection	AI-group scores indistinguishable from random guessing; offline group showed modest genuine knowledge; median AI content 100%	Sample confined to one operations research course; difficult to replicate
Mumtaz [21] (2025)	Multimodal GenAI for personalised learning and adaptive assessment	SLR (2023-2025) plus prototype evaluation; n=30 students; physics teaching	Accuracy improved 65% to 82%; engagement 3.2 to 4.5/5; task completion reduced by 4 min; adaptive assessment supported	Only 30 students; cross-modal coherence and computational demands remain unresolved
Lodzikowski et al. [6] (2023)	Educational implications of GenAI; evolution from rule-based AI to LLMs	Conceptual review with illustrative examples	GenAI enables personalised, adaptive, multimodal learning; human-in-the-loop oversight essential	Empirical validation absent; rapid change may date conclusions
Batista, Mesquita & Carnaz [12] (2024)	SLR of GAI in HE 2023-2024: technology use, acceptance, academic integrity	PRISMA SLR; Scopus and WoS; 37 empirical articles	GAI enhances student support, teaching efficiency and research; assessment integrity primary concern; institutional policy urgently needed	Timeframe limited to Jan 2024; restricted to higher education
Purwono et al. [7] (2025)	GenAI as pedagogical co-pilot in HE: adaptive learning and teacher support	Qualitative thematic analysis; 54 studies from Scopus, ScienceDirect, Google Scholar (2019-2025)	GenAI reduces administrative burden, supports personalised feedback, transforms teacher roles; hallucinations, equity gaps identified	Relies on secondary literature; no empirical validation; not fully PRISMA-compliant
Palaniappan et al. [17] (2025)	AI-driven personalised learning pathways; empirical platform evaluation	System development; K-means clustering; Random Forest; 500 students, 5,000 interactions; synthetic dataset	11.7% mean performance improvement; 6.3% completion rate increase; 78.5% prediction accuracy; 4 learner clusters identified	Synthetic dataset; single institution; no longitudinal sustainability data
Ng & Ho [3] (2025)	Bibliometric mapping of GenAI in education research 2022-2025	Bibliometric analysis; 3,808 Scopus articles; VOSviewer co-occurrence and co-citation	Exponential growth (6 in 2022 to 1,951 in 2024); 6 thematic clusters; research shifting to learner-centred concerns	Scopus only; English-language bias; excludes conference papers
Alfarwan [18] (2025)	GenAI use in K-12 education	Systematic review; PRISMA; Scopus and WoS; n=30 papers (2023-2025)	ChatGPT dominates; STEM focus; quasi-experimental designs most common; high school most studied	English-only corpus; no African or Middle Eastern studies; self-report measures prevalent
Lodge, de Barba & Broadbent [29] (2023)	Self-regulated and co-regulated learning with GenAI in HE	Commentary; theoretical synthesis drawing on SRL, co-regulation, actor-network theory	AI literacy insufficient alone; SRL and co-regulation with GenAI essential; individual agency must remain central	No empirical data; theoretical conclusions require longitudinal validation
Marzano [19] (2025)	GAI in K-12 teaching and learning:	Systematic review; PRISMA with Mentefacto	GAI personalises learning, motivates students, improves	No primary data; European studies under-represented;

	opportunities, challenges, gaps	Map; 5 databases; n=197 studies (2016-2024)	assessment; STEM dominates; teacher training critically lacking	longitudinal evidence absent
Izquierdo-Condoy et al. [30] (2025)	GenAI impact on critical thinking and cognitive autonomy in medical education	Viewpoint; critical synthesis of peer-reviewed studies from PubMed, Scopus, WoS	Structured GenAI use (LearnGuide RCT) improves critical thinking; unsupervised use correlates negatively with reasoning; cognitive offloading as key risk	No systematic search protocol; heterogeneous populations; primary data absent
Bernstein et al. [15] (2025)	Harms and unintended consequences of GenAI in computing education	PRISMA SLR; deductive coding; 224 papers from ACM DL, IEEE Xplore, Scopus (2022-2025)	6 harm categories identified; cognitive harms most prevalent (32.4%); only 21% of 2024 papers discussed harms; chat models most associated with harms	Scoping review; does not measure magnitude of harms; omits K-12
Altares-Lopez et al. [31] (2024)	Accelerated GenAI learning methodology; student perceptions	Pilot study; Likert-scale survey; n=10 high school students; binary questions; Wilcoxon testing	High satisfaction (mean 4.7/5); strong correlation (0.95) between AI topic interest and satisfaction; 90% reported teachers do not encourage AI tool use	Very small sample (n=10); 90% male; single session; limited geographic scope (Madrid)
Bouguettaya et al. [4] (2025)	Meta-survey of GenAI in education: trends, challenges, frameworks	Tertiary systematic review (meta-survey); PRISMA; 21 high-quality secondary studies analysed	Higher education dominates; K-12 under-represented; only 7 of 21 reviews employed formal synthesis; commercial tools increasingly embedded	Meta-survey; cannot validate primary data; English-only literature
Roe & Perkins [32] (2024)	GenAI and agency in education: Critical Digital Pedagogy perspective	PRISMA-ScR scoping review; hybrid thematic analysis; Claude 3.5 Sonnet for triangulation; 11 studies from 4 databases	GenAI may enhance learner agency through personalisation but risks exacerbating inequalities; all 11 studies from 2024	Very small corpus (n=11); dominance of theoretical over empirical papers; GenAI tool in analysis may introduce bias
Lin et al. [33] (2025)	Applications, challenges and prospects of GAI in medical education	Scoping review (PRISMA-ScR); 131 articles from PubMed, WoS, Scopus (Jan 2023-Oct 2024)	74% research from very-high-HDI countries; ChatGPT studied in 89.5% of articles; GAI supports diversified teaching; RMA framework proposed	Search capped at Oct 2024; English-language bias; nursing education excluded
Qian [34] (2025)	Pedagogical applications of GenAI in higher education	Systematic review (PRISMA); 37 peer-reviewed empirical studies from ERIC, WoS, ScienceDirect (2023-2024)	China and US lead publications; ChatGPT dominates; three pedagogical themes: automated feedback, learning support, critical skills; overreliance threatens critical thinking	Geographic concentration; perception-based studies excluded; conference papers excluded
Nhan et al. [35] (2025)	Opportunities and challenges of AI tools in EFL writing instruction	Mixed methods; structured survey (n=150) and semi-structured interviews (n=15); descriptive statistics + thematic analysis	AI tools positively perceived for personalised learning (mean 4.10-4.32); instant feedback rated highly; over-reliance a moderate concern	Single institution (Mekong Delta); predominantly female cohort; no longitudinal tracking
Li [22] (2024)	Usability of ChatGPT in second language acquisition	Systematic literature review; qualitative synthesis of empirical and conceptual studies	ChatGPT NLP capabilities support personalised language learning; improvements in grammar, vocabulary, writing coherence; learning autonomy and motivation increase	Literature review limits causal claims; limited evidence on listening and speaking skills; no longitudinal data
Jin et al. [36] (2024)	Global institutional policies and guidelines for GAI adoption across 40 universities	Qualitative thematic analysis; stratified sampling from QS World Rankings 2024; Cohen's Kappa (all themes >= 0.61)	All 40 universities addressed academic integrity; 38 emphasised teaching/learning enhancement; 33 focused on AI literacy; only 7 engaged in observability measures	Sample bias toward well-resourced institutions; temporal cutoff Jan 2024; focus on official documents

Pinho, Costa & Pinho [37] (2025)	Generative AI governance model in educational research	Scoping review; bibliometric analysis (VOSviewer); content analysis (webQDA); WoS (n=194, 35 selected)	Living GenAI Governance Model at macro, meso and micro levels; HE and LLMs as motor themes; AI literacy and ethics emerging themes	Single database (WoS); English-language bias; governance model not yet empirically validated
Arar et al. [14] (2025)	AI, GenAI and research integrity: hybrid systemic review	Hybrid integrative systematic review; bibliometric analysis; WoS; n=595 English peer-reviewed articles (2010-2025)	Publication volume peaked in 2023-2024 driven by LLMs; key risks: hallucination, algorithmic bias, plagiarism, academic dishonesty, skill erosion	WoS only; English articles exclusively; 6.8% retention rate raises coverage concerns
Garzón, Patiño & Marulanda [38] (2025)	Systematic review of AI in education: trends, benefits, challenges	Systematic review following PRISMA and Kitchenham-Charters; 3 databases; n=155 empirical studies (2015-2025); Cohen's kappa = 0.89	GenAI most implemented AI type (30.3%); HE dominated (62.6%); learning gains in 65.8%, motivation in 30.3%; ethical concerns top challenge (14.8%)	Grey literature excluded; short-term interventions dominate; no evidence certainty framework applied
Sanz-Tejeda et al. [13] (2026)	Impact of GenAI on academic reading and writing in HE social sciences (2023-2025)	Systematic review using SALSA and PRISMA; Scopus and WoS; n=136 articles; NotebookLM for preliminary thematic analysis	Clear trend toward ChatGPT in academic writing; improvements in text coherence, lexical richness, argumentation; only 19-23% assessors correctly identified AI scripts	Restricted to social sciences; Spanish and English only; ERIC and PsycINFO not consulted
Ovi, Fierro & Smith [20] (2025)	Engineering student adoption of GenAI 2023-2024	Longitudinal representative survey; two waves at Colorado School of Mines (May 2023, n=601; Sep 2024, n=862); chi-square, logistic regression	LLM adoption rose significantly; 45.1% regular/daily users by 2024; primary motivations: depth of knowledge (57.7%), quality of work (40.1%); top concern: misinformation (73%)	Single university site; opt-in bias; surveys at different semester points; self-report risk
Chakraborty [1] (2024)	Role of GenAI across schooling, HE and R&D (Education 1.0 to 5.0)	Narrative literature review; multi-stakeholder conceptual framework	GenAI enhances personalised learning, automated grading, teacher professional development; ChatGPT correlated 0.86 with human grades in essay scoring	No primary data collection; lacks longitudinal evidence; limited cultural generalisability
Li et al. [8] (2024)	Integration of GenAI with adaptive learning systems	Position paper with systematic literature review of ML and GenAI works on adaptive learning	GenAI empowers AL via dynamic content creation, intelligent agents, learning simulation; LLMs resolve cold-start challenges in knowledge tracing	Scope limited to ML-related AL components; no empirical validation
Haroud & Saqri [39] (2025)	Teachers' and students' perceptions of GenAI in Moroccan HE	Mixed methods; quantitative survey (130 teachers, 156 students); t-tests; qualitative content analysis	Students scored significantly higher on AI adoption ($p<0.001$); both groups reject teacher replacement; digital literacy identified as critical prerequisite	Single-institution convenience sample; social desirability bias; no longitudinal follow-up
Ogunleye et al. [5] (2024)	Systematic review of GenAI for teaching and learning practice in HE	PRISMA-guided SLR; bibliometric indicators (VOSviewer); Latent Dirichlet Allocation topic modelling; 355 Scopus papers (2017-2023); Cohen's kappa=0.659	Publications grew from 38 (2022) to 273 (2023) post-ChatGPT; 10 core research themes; USA, Australia, China and UK lead authorship	English-language only; Scopus only; no studies beyond 2023
Deng et al. [2] (2025)	Meta-analysis: does ChatGPT enhance student learning?	Systematic review and meta-analysis; PRISMA; 69 experimental and quasi-experimental studies; Hedges' g;	Large positive effect on academic performance ($g+=0.712$), affective-motivational states ($g+=0.881$), higher-order thinking ($g+=0.703$); reduced mental effort ($g+=-0.675$)	Only 8.06% conducted power analysis; non-English studies excluded; restricted to ChatGPT

		random-effects model; trim-and-fill analyses		
Mimoudi [11] (2025)	GenAI opportunities and challenges through an equity lens	Systematic review; PRISMA; 75 peer-reviewed studies (2016-2024); binary thematic coding; Scopus, WoS, Google Scholar	Personalised learning in 60% of studies; data privacy cited in 56%; algorithmic bias in 52%; digital divide in 48%; equity depends on personalisation, fairness and access	No inter-rater reliability; English and French publications only; grey literature excluded
Wu, Dang & Li [9] (2025)	Students' and teachers' responses, attitudes and behaviours toward GenAI in HE	Systematic review; PRISMA; 99 empirical studies (2020-Aug 2024); two-cycle coding; Cohen's Kappa 92%; Scopus, WoS, EBSCO	73 of 99 studies reported positive emotional responses; 52 showed GenAI promoted cognitive processing; 34 studies raised concern over over-reliance	English-language only; restricted to HE; longitudinal data absent; teacher perspectives underrepresented
Wang, Zainuddin & Chin [10] (2025)	Empirical landscape of GenAI in pedagogical practices (2022-2024)	Systematic review; PRISMA; 28 empirical studies (2022-2024); WoS (SSCI-indexed); two independent coders (>90% agreement)	4 positive effects identified; ChatGPT platform in 85.71% of studies; HE accounted for 71.42% of levels investigated; 3 challenges: inaccuracy, academic integrity, inconsistent efficacy	Small sample (28 studies); single database; no grey literature; temporal scope underrepresents early 2022

Several cross-cutting observations emerge from the synthesis. First, ChatGPT and its GPT-lineage variants receive disproportionate research attention relative to other tools, with Alfarwan [18] and Wang et al. [10] finding ChatGPT was the platform in 85.71% of empirical studies.

Applications of Generative AI in Education

- **AI-Assisted Learning and Tutoring**

AI tutoring is the most documented GenAI application. Deng et al. [2] found large positive effects on academic performance (g+ = 0.712) and motivation (g+ = 0.881) across 69 studies, with 79.71% using ChatGPT as a direct learning tool. Classroom-based interventions lasting over one week produced the most consistent results [2].

Khanmigo delivers Socratic guidance by providing hints rather than direct answers, with real-time analytics for educators [7]. Palaniappan et al. [17] found an 11.7% mean performance improvement across 500 students using personalised pathways, with the largest gains among struggling and quick learners.
Personalised learning was the most cited opportunity, appearing in 60% of 75 studies [11]. Students particularly value being able to ask questions without social anxiety, with AI adoption scores significantly higher than teacher perceptions ($t(230.85) = -8.15$, $p < 0.001$) [39].

Tutoring benefits are unevenly distributed. High-performing students improved with Copilot while struggling students did worse [15]. Students who offloaded assignments entirely failed to develop the reflective skills linked to genuine learning [32].

- **Automated Educational Content Creation**

GenAI automates the production of lesson plans, rubrics, MCQs and assessment tasks. MagicSchool AI provides over 50 templates for common instructional needs [7]. Eight studies reported reduced workload and improved efficiency for teachers [9].

In assessment item generation, Claude produced the most error-free items (12 of 20) in a comparative study, though expert validation remained essential in all AI conditions [30]. Lin et al. [33] found 62.5% of studies confirmed AI-generated MCQs were valid. Expert review of AI-generated content remains necessary [12].

Agentic systems can now generate entire course modules in minutes, raising concerns about assessment integrity and requiring governance strategies such as process logging and oral defences [4].

AI-Generated Multimedia Learning Materials

Altares-Lopez et al. [31] found secondary students using Stable Diffusion to create generative videos developed practical prompt engineering skills and achieved high satisfaction (4.7/5). A correlation of 0.88 was found between class participation and cross-technology learning [31].

Mumtaz [21] prototyped a multimodal physics system where AI processed student-drawn diagrams and returned annotated overlays; accuracy improved from 65% to 82% over a text-only group, with task completion time reduced by four minutes. Interactive content achieves the highest engagement scores (0.65) compared to static formats [17].

DALL-E images of retinal disease and skin conditions have been used in medical education, though accuracy remains limited [33]. Image models like DALL-E and Midjourney have been deployed in CS ethics modules to help students identify representational bias, and in STEAM curricula to broaden participation [15][10].

- **AI-Generated Audio and Music for Learning**

Empirical evidence for GenAI in audio or music-based learning is limited. Text-to-speech is used for audio narration and accessibility, whilst speech-to-text supports lecture transcription [19].

Tools such as Suno for music generation exist, but 80.5% of engineering students had never used audio generators [20]. Voice assistants provide audio-mediated support for learners with disabilities [11], and speech recognition reduces cognitive overload in language learning [35].

Audio and music-based GenAI learning is a genuine research gap, with particular relevance for language acquisition, reading difficulties and music education [19][34][11].

- **Research and Academic Assistance**

GenAI supports literature review, summarisation and academic writing. Bouguettaya et al. [4] distinguish retrieval-augmented tools like Perplexity from document-grounded tutors like NotebookLM. The number of arXiv papers referencing ChatGPT grew from 25 in January 2023 to 772 by June 2023 [33].

ChatGPT-assisted writing improved text coherence, lexical richness and argumentation [13]. It provides more consistent structural feedback than peer reviewers, though human peers better identify content problems, suggesting a complementary role [13].

The risks are serious. Arar et al. [14] identify hallucination as a systemic threat to research integrity. ChatGPT-4o generates non-existent references in 46% of cases [13]. Students also confirmed information was often outdated or incorrect [9].

- **Comparative Analysis of Generative AI Tools**

Table 2 maps the principal GenAI tools to their capabilities, educational applications and limitations, based exclusively on the reviewed sources.

Table 2: Tool Capability and Risk Matrix — Generative AI in Education

Tool	Type of Content Generated	Educational Use	Advantages	Limitations
ChatGPT (GPT-3.5/4/4o) [7][3][15][4][33][34][2][9][10]	Text, code, feedback, essays, quiz items, lesson plans, summaries	AI tutoring, personalised feedback, content creation, writing support, exam preparation, language learning, programming assistance	Widely accessible; versatile across disciplines; supports differentiated instruction; multilingual; reduces administrative burden [7][2]	Hallucinations (confident but false content); enables academic dishonesty [2][10]; over-reliance reduces metacognitive engagement [15][9]; outputs can be outdated or unreliable [9]
GitHub Copilot [15][16]	Code completions and suggestions	Code generation and assistance in computing education; AI-first curricula	Reduces syntax burden; enables AI-first curricula [15]	Metacognitive harm; lower-performing students accepted more suggestions and had poorer outcomes; positively correlated with time to task completion (r=0.693, p=0.0005) [15]
Claude (Anthropic) [15][32][33]	Text generation; analytical coding support; long-form reasoning	Qualitative data analysis triangulation; conversational tutoring; assessment item generation	Highest proportion of error-free assessment items (12/20) in comparative study [32]; used for thematic triangulation in research [33]	May introduce biases in analytical processes; limited K-12 evidence; primarily evaluated in medical/research contexts [33]
Google Gemini (DeepMind) [7][3][33][34]	Text, multimodal content (text + image)	General-purpose tutoring; search-integrated learning support; short-answer grading	Strong contextual understanding; Gemini 1.0 Pro short-answer grading showed no significant difference from	Requires Workspace access; inconsistent pedagogical alignment; lower research adoption than ChatGPT (16.5% of articles) [33]

			human evaluators [33]	
DALL-E / Midjourney [18][15][33][11][10]	AI-generated images from text prompts	Visual content creation; exploring representational bias; art and STEAM education	Enables creative exploration; used in CS ethics modules to examine bias [15]; best average score (3.2/5) among image tools for skin ulcer generation [33]	Representational harms (stereotypes); DALL-E overall clinical accuracy 22.2%; 65% of wound images had anatomical inaccuracies [33]; copyright concerns
Stable Diffusion [33][10]	Generative images; synthetic clinical imagery	Medical training (synthesising retinal disease images); STEAM education	Enhances availability of visual teaching resources; open-source accessibility [33]	Quality varies significantly; expert clinical review required before classroom use [33]
Deforum Stable Diffusion [31]	Generative video from sequential text prompts	Hands-on AI video creation; prompt engineering practice	Open-source; stimulates creativity and interdisciplinary investigation; facilitated spontaneous ethical discussion on representational bias [31]	Output does not always match student intent; frustration reported [31]; very small pilot (n=10)
Khanmigo (Khan Academy) [7][4]	Socratic text guidance; personalised feedback	One-to-one tutoring in mathematics and science; curriculum-aligned guided prompts	Pedagogically aligned; provides hints rather than direct answers; real-time classroom analytics [7]	Limited to Khan Academy ecosystem; constrained scalability outside the platform [7]
MagicSchool AI [7]	Lesson plans, worksheets, rubrics, parent communications	Teacher support: automated planning, IEP drafts, formative question generation	50+ AI templates for common instructional tasks; reduces administrative burden [7]	Requires account verification; subscription tiers limit access; quality of generated content varies [7]
Diffit [7]	Adapted reading passages; level-appropriate quizzes	Content differentiation for diverse literacy levels; multilingual support	Level-based Lexile output; accessible multilingual scaffolding [7]	Best suited to language arts; limited STEM content [7]
Multimodal GenAI Prototype [21]	Text, speech, image and interactive simulation combined	Adaptive assessment; personalised physics tutoring; collaborative learning facilitation	Accuracy improved 65% to 82%; engagement 3.2 to 4.5/5; task completion reduced by 4 min [21]	Prototype scale (n=30); computationally intensive; not commercially available at scale [21]
LearnLM (Gemini-based) [1]	Conversational tutoring text; interactive task generation	One-to-one conversational tutoring aligned to individual learning style	Adapts to student demands in real time; built for educational context [1]	Limited deployment evidence; dependent on quality of instructional prompts [1]
LLaMA (Meta AI) [33]	Text: question answering, explanations	Low-resource educational environments requiring local model deployment	Runs locally; suitable for environments with limited internet resources [33]	Referenced in only 4 articles [33]; limited empirical evaluation in educational settings
Customised GPT / RAG Chatbots	Discipline-specific explanations;	Anatomy education, biology module Q&A, EFL	Tailored capabilities overcome limitations of general-purpose	RAG chatbots restrict broader application; 50% of

(AnatomyGPT, Q-Module-Bot) [33][34]	immediate module-specific Q&A	personalised writing exercises	LLMs; AnatomyGPT provides detailed feedback with rationales and citations [33]	users found Q-Module-Bot non-intuitive [34]
Grammarly / QuillBot / Wordtune [35][22][13]	Grammar corrections; paraphrased text; vocabulary suggestions	Writing improvement; grammar and spelling correction; EFL classrooms	Positive impact on grammatical accuracy and lexical richness; reduces foreign language anxiety [22]	Primarily surface-level corrections; does not develop higher-order writing skills [35]; over-reliance risk
Eduaide.ai [27]	Instructional resources; quiz generators; rubrics aligned to Bloom's Taxonomy	Assessment design aligned with cognitive levels; differentiated task generation	Over 100 instructional resource formats; explicit Bloom's taxonomy alignment; freemium access [27]	Up to 16,000 token input limit; primarily instructor-facing; limited student interaction [27]
Perplexity [4]	Retrieval-augmented text with live web citations	Research assistance; source evaluation practice	Real-time source grounding; citation trails support academic workflows [4]	Risk of uncritical reliance on automatically retrieved content [4]
Suno AI [34]	AI-generated music content	Music generation as part of broader GenAI ecosystem	Single-modal creative content generation; identified as part of emerging GenAI landscape [34]	No empirical study in reviewed corpus investigated audio/music-based GenAI pedagogy [34]

Accessibility remains a cross-cutting inequality. Differential benefits consistently favour already well-prepared students [15]. Unequal access was identified in 48% of reviewed studies, with personalised learning conditional on connectivity, devices and institutional readiness [11].

Benefits of Generative AI in Education

The reviewed literature documents benefits across four domains.

Personalised learning is the most validated benefit. Deng et al. [2] found large effects on performance (g+ = 0.712) and motivation (g+ = 0.881). Learning gains appeared in 65.8% of 155 empirical studies [38]. AI scaffolding that reduces support as learner competence grows can build genuine independence [7].

Instructional efficiency is a second benefit. Eight studies reported GenAI reduced administrative and grading burden [9]. MagicSchool AI provides over 50 templates for common instructional tasks, and ChatGPT-generated course goals achieved 92% expert ratings for appropriateness [33].

Accessibility and inclusivity represent a third benefit. Expanded access was identified in 44% of studies, with text-to-speech, subtitles and real-time translation supporting diverse learning needs [11]. Free-tier access was used by 92.9% of students, democratising baseline GenAI use [20].

Enhanced engagement and motivation is a fourth benefit. Wu et al. [9] found 73 of 99 studies reported positive emotional responses. In engineering, 57.7% of students cited deepening understanding as their primary motivation for LLM use [20], suggesting genuine rather than shortcut-oriented engagement.

Challenges and Ethical Issues

The reviewed literature also identifies significant challenges requiring institutional governance and continued research.

Academic integrity is the most consistently documented concern. Only 23% of evaluators correctly identified AI-generated text [13]. In Benedek and Sziklai's [28] trial, students with AI access produced offline test scores indistinguishable from random guessing, indicating knowledge levels below 20%. Apparent learning gains in 42 of 51 studies may partly reflect AI-generated content rather than genuine knowledge [2].

Hallucination poses a serious technical risk. ChatGPT-4o generates non-existent references in 46% of cases [13]. Arar et al. [14] characterise hallucination as a systemic property rather than an isolated defect, and 44 of 99 studies raised concerns about content quality [9].

Algorithmic bias was identified in 52% of reviewed studies [11]. Altares-Lopez et al. [31] observed direct representational bias in a classroom activity, where the AI failed to depict Black individuals in higher socioeconomic positions in historical frames, triggering an in-session ethical discussion. Image models have since been used in CS ethics modules to teach students to identify such bias [15].

Over-reliance on AI tools and erosion of critical thinking and cognitive autonomy represent a fourth well-documented concern. Izquierdo-Condoy et al. [30] found that unsupervised ChatGPT use correlates negatively with reasoning scores, and identifies cognitive offloading as a key risk whereby learners delegate analytical tasks to the AI, eroding their own reasoning structures. Rajesh Kumar et al. [16] report that the unguided use of AI tools is consistently linked to decreased academic performance, including lower GPAs and overall exam scores. Wu, Dang and Li [9] found that 34 studies raised concern over over-reliance on generated content, and Ovi et al. [20] found that 20.8% of engineering students reported generating solutions without first solving problems themselves.

Data privacy was cited as a challenge in 56% of studies [11]. Multimodal systems collect particularly sensitive data including handwriting, speech and gaze patterns, raising acute data governance concerns [21].

VI. RESEARCH GAPS AND FUTURE RESEARCH DIRECTIONS

The reviewed literature identifies several under-investigated areas.
Longitudinal evidence is the most universally missing element. All five core reviewed studies flag it as a critical limitation [6][12][7][17][3]. Short-term interventions cannot assess whether GenAI-supported learning produces durable gains or cumulative skill degradation over time, particularly given concerns about cognitive offloading [30][28][16].

Teacher training and professional development for AI integration represents a second critical gap. Marzano [19] identifies teacher training and ministerial guidelines as critically lacking, with 197 reviewed studies revealing an inadequate empirical evidence base for effective professional development models that equip educators to deploy GenAI tools responsibly in classroom contexts. Dhakal [24] corroborates this finding through bibliometric analysis, documenting that educator professional development is one of the most urgent needs identified across 310 research outputs on the GenAI-pedagogy nexus. Raman et al. [23] note that institutional readiness is a key antecedent to effective GenAI adoption, and that without structured professional development frameworks, the transformative potential of GenAI risks being undermined by inconsistent and unsupported implementation.
Policy frameworks remain underdeveloped. Fewer than half of top universities had public GenAI guidelines, and over 94% of educators reported no clear institutional policy [25]. Only 7 of 40 universities were conducting formal evaluation of GenAI impact [36]. Accountability gaps when AI systems cause harm remain largely unaddressed [14].

Global South representation is a major structural gap. The US (935), China (508) and Hong Kong (147) dominate output [3]. Structural barriers including research funding, infrastructure and English-dominant publication norms account for this disparity rather than lack of interest [11]. This undermines the generalisability of current findings [38].

Audio and music-based GenAI learning is conspicuously absent from the reviewed literature [19][34][20]. As voice-based interfaces proliferate, adaptive spoken explanations and audio feedback become increasingly important for accessibility and language acquisition. This is a clear priority area for future research.

K-12 evidence remains insufficient. All 30 empirical K-12 studies were published after 2023 [18], and early childhood is barely represented. Bernstein et al. [15] excluded K-12 from their harm review due to insufficient evidence. Given the societal significance of foundational schooling, this gap carries major policy implications.

VII. CONCLUSION

This review synthesises 39 peer-reviewed studies from 2022 to 2026 on GenAI in education. The central finding is that GenAI delivers real and substantial benefits in personalised learning and content creation, but its safe realisation depends on institutional governance and pedagogical scaffolding that most systems currently lack.

Structured, educator-mediated integration produces meaningful learning gains. Unguided use introduces academic integrity risks, cognitive dependency and equity amplification. Because AI detection tools can be evaded [12], institutions must move toward assessment designs intrinsically resistant to AI substitution, such as oral defences, process portfolios and collaborative live tasks. Educators also require professional development to mitigate cognitive offloading and foster co-regulated learning [4].

The field faces urgent structural gaps. The severe underrepresentation of K-12 environments [18][19], the Global South [11][3], and audio-based learning contexts [19][34] severely limits the global generalisability of current findings. Most critically, the absence of longitudinal evidence on the long-term cognitive effects of AI-assisted learning remains the most universally identified limitation across the literature. The 2022–2026 period has established the factual basis for GenAI's educational potential; the defining challenge of the next research phase is building the institutional and governance infrastructure to realise that potential equitably, safely, and in ways that strengthen the independent cognitive development at the heart of education.

REFERENCES

1. S. Chakraborty, "Generative AI in Modern Education Society," arXiv:2412.08666v1 [cs.CY], Dec. 2024.
2. R. Deng, M. Jiang, X. Yu, Y. Lu, and S. Liu, "Does ChatGPT enhance student learning? A systematic review and meta-analysis of experimental studies," Computers & Education, vol. 227, p. 105224, 2025, doi: 10.1016/j.compedu.2024.105224.
3. S.-L. Ng and C.-C. Ho, "Generative AI in education: Mapping the research landscape through bibliometric analysis," Information, vol. 16, no. 8, p. 657, Jul. 2025, doi: 10.3390/info16080657.
4. S. Bouguettaya, F. Pupo, M. Chen, and G. Fortino, "A Meta-Survey of Generative AI in Education: Trends, Challenges, and Research Directions," Big Data and Cognitive Computing, vol. 9, no. 237, 2025, doi: 10.3390/bdcc9090237.
5. B. Ogunleye, K. I. Zakariyyah, O. Ajao, O. Olayinka, and H. Sharma, "A Systematic Review of Generative AI for Teaching and Learning Practice," Educ. Sci., vol. 14, no. 6, p. 636, Jun. 2024, doi: 10.3390/educsci14060636.
6. K. Lodzikowski, P. W. Foltz, and J. T. Behrens, "Generative AI and its educational implications," in Advances in Education Research, 2023.
7. Purwono, A. Ma'arif, H. M. Marhoon, A. J. Abougarair, and A.-N. Sharkawy, "Generative AI as a pedagogical co-pilot in higher education: A review of adaptive learning and teacher support systems," Journal of Technological Pedagogy and Educational Development, vol. 2, no. 1, pp. 28-42, 2025, doi: 10.59247/jtped.v2i1.8.
8. H. Li, T. Xu, C. Zhang, E. Chen, J. Liang, X. Fan, H. Li, J. Tang, and Q. Wen, "Bringing Generative AI to Adaptive Learning in Education," arXiv:2402.14601v3 [cs.CY], Jun. 2024.
9. F. Wu, Y. Dang, and M. Li, "A systematic review of responses, attitudes, and utilization behaviors on generative AI for teaching and learning in higher education," Behavioral Sciences, vol. 15, no. 4, p. 467, 2025, doi: 10.3390/bs15040467.
10. X. Wang, Z. Zainuddin, and H. L. Chin, "Generative artificial intelligence in pedagogical practices: a systematic review of empirical studies (2022-2024)," Cogent Education, vol. 12, no. 1, p. 2485499, 2025, doi: 10.1080/2331186X.2025.2485499.
11. A. Mimoudi, "Generative AI to bridge the educational divide: Personalized learning and challenges," Social Sciences & Humanities Open, vol. 12, p. 102140, 2025, doi: 10.1016/j.ssaho.2025.102140.
12. . Batista, A. Mesquita, and G. Carnaz, "Generative AI and higher education: Trends, challenges, and future directions from a systematic literature review," Information, vol. 15, no. 11, p. 676, Oct. 2024, doi: 10.3390/info15110676.
13. A. Sanz-Tejeda, J. C. Domínguez-Oller, J. M. Baldaquí-Escandell, R. Gómez-Díaz, and A. García-Rodríguez, "The impact of generative AI on academic reading and writing: a synthesis of recent evidence (2023-2025)," Front. Educ., vol. 10, p. 1711718, Jan. 2026, doi: 10.3389/feduc.2025.1711718.
14. K. H. Arar, H. Özen, G. Polat, and S. Turan, "Artificial intelligence, generative artificial intelligence and research integrity: a hybrid systemic review," Smart Learn. Environ., vol. 12, p. 44, 2025, doi: 10.1186/s40561-025-00403-3.
15. S. Bernstein, A. Rahman, N. Sharifi, A. Terbish, and S. MacNeil, "Beyond the Benefits: A Systematic Review of the Harms and Consequences of Generative AI in Computing Education," in Proc. 25th Koli Calling Int. Conf. Computing Education Research, Koli, Finland, 2025, doi: 10.1145/3769994.3770036.
16. P. Rajesh Kumar, S. Md. Yusof, R. Che Aziz, and A. Mahmud, "The Impact of Generative AI Tools on Academic Performance in Higher Education: A Systematic Review (2023-2025)," Int. J. Acad. Res. Bus. Soc. Sci., vol. 15, no. 12, pp. 2149-2160, Dec. 2025, doi: 10.6007/IJARBSS/v15-i12/27076.
17. S. Palaniappan, K. Subaramaniam, L. T. K. Liew, R. Logeswaran, and O. Baker, "AI-driven personalized learning pathways: Transforming educational outcomes through adaptive content delivery systems," International Journal of Modern Education, vol. 7, no. 27, pp. 692-705, Oct. 2025, doi: 10.35631/IJMOE.727043.
18. A. Alfarwan, "Generative AI use in K-12 education: a systematic review," Frontiers in Education, vol. 10, p. 1647573, 2025, doi: 10.3389/feduc.2025.1647573.
19. D. Marzano, "Generative artificial intelligence (GAI) in teaching and learning processes at the K-12 level: a systematic review," Technology, Knowledge and Learning, 2025, doi: 10.1007/s10758-025-09853-7.
20. J. A. Ovi, G. Fierro, and C. E. Smith, "Assessing Student Adoption of Generative Artificial Intelligence across Engineering Education from 2023 to 2024," arXiv:2503.04696, Mar. 2025.
21. F. Mumtaz, "Multimodal Generative AI for Education and Learning Systems," School of Computing, Engineering and Built Environment, Edinburgh Napier Univ., 2025.
22. . Li, "Usability of ChatGPT in second language acquisition: Capabilities, effectiveness, applications, challenges, and solutions," Studies in Applied Linguistics & TESOL, vol. 24, no. 1, pp. 24-37, 2024.
23. R. Raman, K. Achuthan, and P. Nedungadi, "Generative AI Integration in Education: Theoretical Review and Future Directions Informed by the ADO Framework," Information, vol. 17, no. 3, p. 241, Mar. 2026, doi: 10.3390/info17030241.
24. S. P. Dhakal, "A Scoping Review of Generative Artificial Intelligence (GenAI) and Pedagogy Nexus: Implications

for the Higher Education Sector," Metrics, vol. 2, no. 3, p. 17, Sep. 2025, doi: 10.3390/metrics2030017.
25. N. J. Francis, S. Jones, and D. P. Smith, "Generative AI in Higher Education: Balancing Innovation and Integrity," British Journal of Biomedical Science, vol. 81, p. 14048, Jan. 2025, doi: 10.3389/bjbs.2024.14048.
26. B. Ogunleye, K. I. Zakariyyah, O. Ajao, O. Olayinka, and H. Sharma, "Higher education assessment practice in the era of generative AI tools," Journal of Applied Learning & Teaching, vol. 7, no. 1, 2024, doi: 10.37074/jalt.2024.7.1.28.
27. G. Ilieva, T. Yankova, M. Ruseva, and S. Kabaivanov, "A Framework for Generative AI-Driven Assessment in Higher Education," Information, vol. 16, no. 6, p. 472, 2025, doi: 10.3390/info16060472.
28. M. Benedek and B. R. Sziklai, "Impact of AI Tools on Learning Outcomes: Decreasing Knowledge and Over-Reliance," arXiv:2510.16019, Oct. 2025.
29. J. M. Lodge, P. de Barba, and J. Broadbent, "Learning with generative artificial intelligence within a network of co-regulation," Journal of University Teaching & Learning Practice, vol. 20, no. 7, 2023, doi: 10.53761/1.20.7.02.
30. J. S. Izquierdo-Condoy, M. Arias-Intriago, A. Tello-De-la-Torre, F. Busch, and E. Ortiz-Prado, "Generative artificial intelligence in medical education: enhancing critical thinking or undermining cognitive autonomy?" Journal of Medical Internet Research, vol. 27, p. e76340, 2025, doi: 10.2196/76340.
31. S. Altares-Lopez, J. M. Bengochea-Guevara, C. Ranz, H. Montes, and A. Ribeiro, "Generative AI: The Power of the New Education," arXiv:2405.13487v1 [cs.CY], May 2024.
32. J. Roe and M. Perkins, "Generative AI and Agency in Education: A Critical Scoping Review and Thematic Analysis," Preprint, Nov. 2024.
33. Y. Lin, Z. Luo, Z. Ye, N. Zhong, L. Zhao, L. Zhang, X. Li, Z. Chen, and Y. Chen, "Applications, challenges, and prospects of generative artificial intelligence empowering medical education: Scoping review," JMIR Med. Educ., vol. 11, e71125, Oct. 2025, doi: 10.2196/71125.
34. Y. Qian, "Pedagogical applications of generative AI in higher education: A systematic review of the field," TechTrends, vol. 69, pp. 1105-1120, Jun. 2025, doi: 10.1007/s11528-025-01100-1.
35. L. K. Nhan, N. T. M. Hoa, and L. V. N. Quang, "Leveraging AI for writing instruction in EFL classrooms: Opportunities and challenges," Educational Process: International Journal, vol. 15, e2025158, 2025, doi: 10.22521/edupij.2025.15.158.
36. Y. Jin, L. Yan, V. Echeverria, D. Gasevic, and R. Martinez-Maldonado, "Generative AI in Higher Education: A Global Perspective of Institutional Adoption Policies and Guidelines," arXiv:2405.11800, May 2024.
37. I. Pinho, A. P. Costa, and C. Pinho, "Generative AI Governance Model in Educational Research," Front. Educ., vol. 10, p. 1594343, Jul. 2025, doi: 10.3389/feduc.2025.1594343.
38. J. Garzón, E. Patiño, and C. Marulanda, "Systematic Review of Artificial Intelligence in Education: Trends, Benefits, and Challenges," Multimodal Technol. Interact., vol. 9, no. 8, p. 84, Aug. 2025, doi: 10.3390/mti9080084.
39. S. Haroud and N. Saqri, "Generative AI in Higher Education: Teachers' and Students' Perspectives on Support, Replacement, and Digital Literacy," Educ. Sci., vol. 15, no. 4, p. 396, Mar. 2025, doi: 10.3390/educsci15040396.