INTERNATIONAL JOURNAL OF EDUCATION AND INFORMATION TECHNOLOGIES
DOI: 10.46300/9109.2025.19.11 Volume 19, 2025

ChatGPT and AI Chatbots in Education: An Umbrella Review of Systematic Reviews, Scoping Reviews, and Meta-Analyses

Emmanouil D. Milakis[1], Constantina Corazon Argyrakou[2], Alexandros Melidis[3], John Vrettaros[4]
[1]European University Cyprus,
6, Diogenes, Engomi, 2404,
Cyprus

[2]Harokopio University of Athens,
70, El. Venizelou, Kallithea, 17676,
Greece

[3]National and Kapodistrian University of Athens, GFOSS - Open Technologies Alliance,
9, Iroon Politechniou, Zografou, 15773,
Greece

[4]GFOSS - Open Technologies Alliance, Ministry of Education, Religious Affairs and Sports,
37, A. Papandreou, Marousi, 15180,
Greece

Received: May 29, 2024. Revised: March 14, 2025. Accepted: May 16, 2025. Published: July 23, 2025.

Abstract—This umbrella review synthesizes findings from 41 systematic reviews, scoping reviews, and meta-analyses on the use of ChatGPT and similar large language model (LLM)-based chatbots in education. It provides a critical analysis of their pedagogical applications, benefits, limitations, and associated ethical and policy issues across diverse educational levels and domains. The evidence reveals that chatbots are predominantly implemented in higher education, with growing use in medical, STEM, and language learning contexts. Reported benefits include personalized learning, enhanced writing and critical thinking skills, and increased learner autonomy. However, significant concerns persist regarding the reliability of AI-generated content, overreliance by students, academic integrity, and institutional preparedness. The review highlights methodological gaps in current research, such as a lack of longitudinal studies and limited attention to underrepresented populations and educational settings. The findings aim to inform evidence-based decision-making for educators, researchers, and policymakers navigating the integration of AI chatbots into formal education systems.

Keywords—ChatGPT, AI Chatbots, Large language models, Education, Educational policy, Umbrella review.

I. Introduction

THE emergence of advanced generative artificial intelligence (AI) technologies, particularly large language models (LLMs), has initiated important shifts in contemporary educational theory and practice [1]. Within this landscape, ChatGPT and other LLM-based chatbots have gained rapid traction, prompting renewed interest in their educational possibilities as well as their limitations [2]. Their broad availability and intuitive design have led educators and institutions to consider their application across diverse contexts, such as academic writing support, language instruction, tutoring, assessment, and student engagement [3].

The rapid development of AI technologies, particularly those built on large-scale neural networks, has led to notable changes in educational practices across multiple levels [4]. These tools are increasingly valued for their capacity to support personalized instruction, generate automated feedback, and foster greater learner autonomy. Nevertheless, the field remains fragmented from both a methodological and conceptual standpoint, with marked inconsistencies in research approaches, outcome metrics, and pedagogical

implementations [5]. An additional challenge lies in the constant release of new chatbot models and updates, which often exceeds the pace at which researchers can rigorously examine their educational impact [6]. As a result, educators, researchers, and policymakers continue to face considerable uncertainty about the pedagogical effectiveness, ethical challenges, and institutional readiness associated with their adoption [7].

Moreover, as educational systems worldwide confront the pedagogical and ethical challenges posed by generative AI, international policy initiatives underscore the need for thoughtful and equitable implementation. Documents such as the UNESCO Guidelines for AI in Education [8], the EU Artificial Intelligence Act [9], and the OECD AI Principles [10] emphasize the importance of integrating chatbots into formal education in ways that promote responsibility, fairness, and transparency.

In response to this evolving and complex research landscape, this umbrella review aims to synthesize evidence from existing systematic reviews, meta-analyses, and meta-syntheses on the educational use of chatbots, in order to provide an integrative understanding of current trends, benefits, limitations, and policy implications. In particular, it addresses the following research questions:

- RQ1: What types and pedagogical functions of chatbots are identified in the literature, and in which educational contexts and subject domains are they applied?
- RQ2: What kinds of learning outcomes are associated with the use of chatbots in education, including cognitive, affective, and behavioral dimensions?
- RQ3: How do students and educators perceive the benefits and limitations of chatbots, and in what ways is the role of the educator evolving in AI-supported learning environments?
- RQ4: What ethical concerns and institutional or policy-level responses are discussed in relation to the integration of chatbots into formal education?

By addressing these questions, this review seeks to clarify the current state of research on educational chatbots and to identify knowledge gaps, methodological limitations, and future directions. The findings are intended to support evidence-informed decision-making for educators, researchers, and policymakers navigating the pedagogical integration of AI-driven tools**.**

II. METHODOLOGY

A systematic literature search was conducted in accordance with the PRISMA 2020 guidelines to identify review studies focusing on the educational applications of LLM chatbots. The search employed the Boolean query (chatbot OR ChatGPT OR Claude OR Gemini OR DeepSeek OR Grok) AND education AND ("literature review" OR review OR "meta-analysis" OR "meta-synthesis") and was carried out between January and May 2025. The process of study identification, screening, eligibility, and inclusion was guided by the PRISMA 2020 flow diagram.

Initially, a total of 351 records were retrieved from three major databases: 232 from Google Scholar, 77 from Scopus, and 42 from Web of Science. A publication year filter was applied to include only studies published between 2023 and 2025. The choice of this publication window was informed by the technological timeline of LLM chatbot deployment in public and educational discourse. ChatGPT was first introduced by OpenAI as a research preview on November 30, 2022, followed by the release of GPT-4 on March 14, 2023. Within the same month, competing systems such as Claude (Anthropic), Gemini (Google Bard), and Grok (xAI) also emerged. Thus, selecting studies from 2023 onward ensured that only literature reflecting this new wave of educational chatbot integration was considered. This resulted in the exclusion of 43 records from Google Scholar and one from Scopus due to earlier publication dates.

At this point, a methodological decision was made to retain only articles indexed in Scopus and Web of Science. This choice was guided by the greater reliability, standardized indexing, and consistent metadata offered by these databases, along with their compatibility with bibliometric analysis tools. Although Google Scholar was initially part of the search process, its entries were excluded from the final analysis because it was difficult to confirm the peer-reviewed status of records and to manage duplicate items. As a result of this filtering process, 189 studies were removed, leaving a final set of 118 records.

A language filter was then applied to retain only publications written in English. As a result, two non-English articles from Scopus and one from Web of Science were excluded. In the next step, only studies that had reached their final publication stage were considered eligible. This led to the removal of two in-press articles from Scopus and four early-access articles from Web of Science. After applying these filters, 70 records from Scopus and 37 from Web of Science remained for full-text screening. In the subsequent phase, records from Scopus and Web of Science were merged and cross-checked. A deduplication process was carried out using manual verification, which resulted in the removal of 23 duplicate entries, reducing the dataset to 86 unique studies.

A detailed full-text review was then conducted to assess each study's relevance to the research scope. To be included, studies had to qualify as literature reviews in the form of systematic reviews, meta-analyses, scoping reviews, bibliometric reviews, narrative reviews, or rapid reviews and had to focus specifically on the use of chatbot technologies in educational contexts. Most of these excluded studies either addressed broader artificial intelligence applications or examined technical aspects unrelated to education or pedagogy. Through this assessment, 45 studies were excluded, resulting in a final set of 41 eligible studies. All records were independently screened by two reviewers. The initial screening of titles and abstracts, as well as the assessment of full-text eligibility, were conducted separately. Any discrepancies

between the reviewers were resolved through discussion until consensus was reached. This dual-reviewer approach was implemented to minimize selection bias and enhance the rigor of the study selection process, as recommended in systematic review methodology. A visual representation of this multi-stage selection process is provided in Figure 1, following the PRISMA 2020 guidelines.

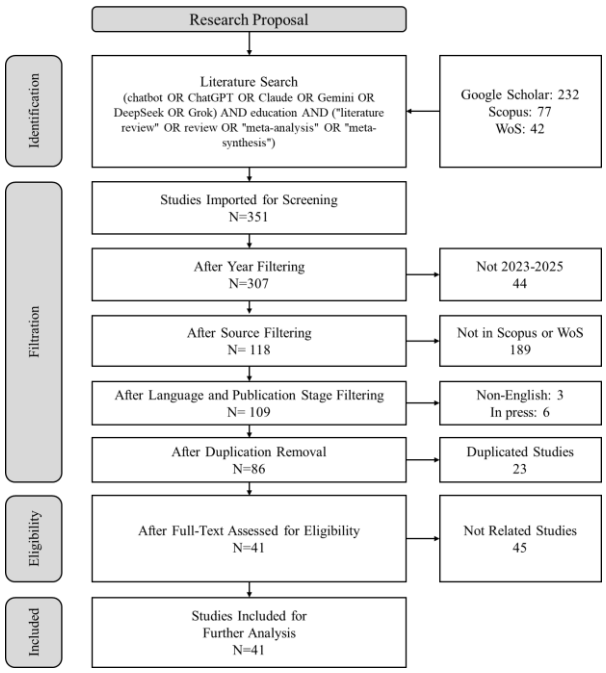

Fig. 1 Flow diagram of study identification and selection

Following the full-text screening process, a final sample of 41 eligible review studies was identified, all of which met the predefined inclusion criteria. These studies constitute the analytical corpus of the present umbrella review. An overview of the 41 included studies is provided in Appendix (Table I). A thematic synthesis was subsequently conducted to explore key dimensions, including the effectiveness of chatbot-assisted learning, levels of student engagement, ethical considerations, the evolving role of the educator, and the development of AI literacy within educational contexts. A thematic synthesis approach was selected to accommodate the heterogeneous nature of the included studies and to identify recurring pedagogical, ethical, and policy-related patterns across diverse educational settings. This method enabled the extraction of cross-cutting themes aligned with the research questions.

Efforts were made to ensure methodological robustness and transparency. Nonetheless, several limitations should be acknowledged. These include the exclusion of non-English studies, the potential for publication bias within indexed databases, and the considerable variation in the quality of the included reviews. Ethical approval was not required, as this umbrella review relied exclusively on secondary analysis of publicly available literature. The review process followed principles of transparency, neutrality, and structured documentation to minimize selection bias.

Although the review applied strict inclusion criteria and followed PRISMA 2020 reporting standards, no formal critical appraisal checklist was used to assess the methodological quality of the included studies. This decision was informed by the diverse nature of the literature, which encompassed systematic, scoping, bibliometric, and narrative reviews. Methodological differences were carefully considered during synthesis and are addressed throughout the interpretation of findings.

III. RESULTS

A. Methodological Characteristics of the Included Reviews

A large proportion of the studies adhered to systematic review protocols, particularly PRISMA [11], [12], [13], with several extending these through scoping [14], [15] or rapid review formats [11], [16]. Bibliometric analyses and hybrid SLR-bibliometric designs were also prominent, employing tools such as VOSviewer and R Studio to map trends and co-authorship networks [17], [18], [19]. Some studies integrated qualitative coding or network analysis techniques [20], while others favored narrative or conceptual syntheses to capture thematic richness [21], [22].

Several reviews focused primarily on conceptual, theoretical, or policy-based discourse [23], [24], whereas others systematically synthesized experimental and quasi-experimental studies, particularly in health and higher education contexts [25], [26]. Reviews that examined ChatGPT specifically tended to rely on early-stage, short-term studies using self-reported or observational data [13], [27], [28], with limited longitudinal evidence or multi-stakeholder involvement. Gaps in faculty, institutional, and cross-cultural perspectives were frequently noted [12], [29].

In addition to the overarching findings synthesized from the reviewed literature, this umbrella review examined the methodological features of the 41 included reviews themselves, revealing distinct patterns in review type and research design. Out of the total sample, 26 studies employed a systematic review design, while 6 were scoping reviews, 5 involved meta-analyses or bibliometric approaches, and 4 followed a narrative review format. In terms of methodological orientation, qualitative designs were the most common, featuring in 24 studies. Mixed methods were employed in 12 reviews, whereas only 5 reviews had a strictly quantitative focus. A visual summary of these distributions is presented in Figure 2, which illustrates the typology and methodological orientation of the included reviews.

A recurrent methodological theme reported within the reviews themselves was the dominance of quantitative designs among the primary studies they analyzed, particularly in evaluating learning outcomes and chatbot efficacy [5], [30]. Nevertheless, qualitative and mixed-methods approaches were increasingly adopted to capture student attitudes, usage patterns, and ethical concerns [31], [32]. While some reviews reported a balanced distribution of methods [13], others emphasized the need for deeper interpretive analysis and

iterative evaluation, especially in dynamic or discipline-specific environments [20], [33].

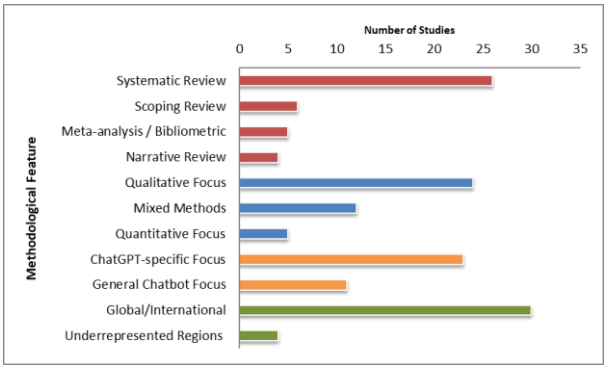

Fig. 2 Methodological features in reviewed studies

Across the evidence base, reviews focusing on ChatGPT (as opposed to chatbots in general) highlighted particular methodological challenges, including the rapid pace of publication, variability in research quality, and the predominance of exploratory designs. Calls for standardization, robust evaluation metrics, and discipline-specific frameworks were common [28], [33]. Several reviews underscored the importance of integrating ethical, pedagogical, and sociotechnical perspectives into future methodologies [34], [35].

Geographically, the reviews exhibited a strong representation from high-income and Anglophone countries, notably the United States, United Kingdom, Australia, and Canada [15], [36]. Some studies offered broader international coverage, but these often lacked contextual specificity or analytical depth regarding regional differences [37], [38]. A few reviews explicitly highlighted underrepresented contexts, such as Iran [39] or the Global South [40], [41], though such cases were exceptions.

B. Educational Contexts of Application

The majority of systematic reviews and meta-analyses converge on tertiary education as the primary site of implementation, often involving undergraduate and postgraduate students across disciplines such as medicine, nursing, engineering, and the humanities [14], [22], [40], [41]. These deployments typically occur in digitally mediated environments such as online, blended, or hybrid settings, where chatbots support asynchronous learning, academic writing, examination preparation, and self-directed study [12], [27], [28].

Beyond higher education, a smaller but growing body of studies addresses secondary education [19], [24], [42]. Primary education remains markedly underrepresented, with only a limited number of reviews identifying empirical applications in this domain [29], [36]. Adult and lifelong learning contexts emerge sporadically across the reviewed literature. These implementations include continuing professional development in healthcare and teacher education, as well as informal learning through MOOCs and other open platforms [26], [43], [44]. Special education is another underexplored area, with only a few studies noting chatbot use in contexts involving learners with diverse needs, despite the technology's theoretical potential for supporting differentiated instruction [5], [45]. A visual summary of educational levels addressed in the reviewed studies is presented in Figure 3.

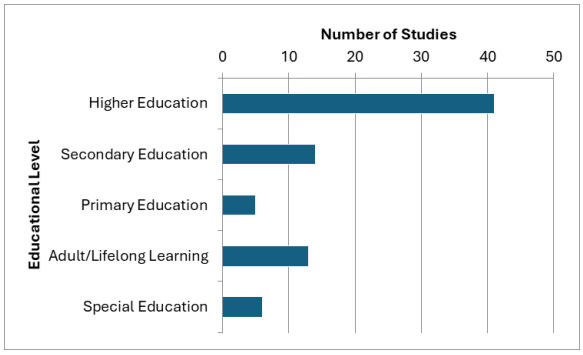

Fig. 3 Educational levels in AI chatbot research

In terms of disciplinary domains, the reviewed studies identify several clusters of chatbot application. A dominant cluster involves health and medical education, where ChatGPT supports clinical reasoning, anatomy learning, and exam preparation in undergraduate, graduate, and residency-level programs [33], [39], [46]. Language learning represents another prominent domain, spanning educational levels from preschool to higher education. In these settings, chatbots act as conversational agents, facilitating vocabulary acquisition, grammar practice, and communicative competence, often in EFL/ESL contexts [13], [32].

STEM education, particularly in engineering and computer science, also figures prominently, with chatbots employed to scaffold problem-solving and programming tasks [19], [31]. ChatGPT, in particular, is used to support academic writing across disciplines, offering automated feedback and enhancing students' capacity for scholarly expression [21], [28]. Interdisciplinary applications are also emerging, particularly in areas such as educational policy, ethics, and digital literacy, often linked to broader educational reform narratives like Education 5.0 [17]. Figure 4 provides a thematic breakdown of the disciplinary domains in which AI chatbots and ChatGPT have been implemented.

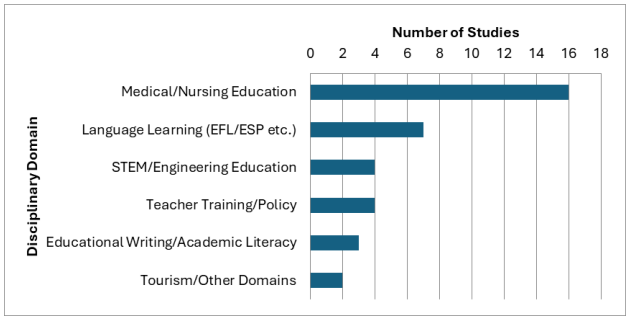

Fig. 4 Disciplinary domains of chatbot/ChatGPT applications

C. Types and Characteristics of Chatbots

The reviewed literature reveals a clear dichotomy in educational chatbot technologies, distinguishing traditional rule-based systems from the more recent generation of AI-powered LLMs, such as ChatGPT. This typological distinction underpins a broader shift in the design, implementation, and pedagogical affordances of chatbots within educational contexts.

Early chatbot implementations in education were primarily rule-based, relying on predetermined scripts, decision trees, and structured input-output sequences [36], [37], [38]. These systems were generally designed for specific instructional tasks, such as responding to frequently asked questions, administering quizzes, or conducting simple tutoring dialogues [25], [47]. Their deterministic structure made the interactions predictable, but also restricted their ability to adapt to different contexts. Most of these chatbots were integrated into messaging platforms or learning management systems (LMS), where they offered basic instructional support [5], [32]. Although they helped automate repetitive teaching duties, their contribution to meaningful pedagogical engagement was limited. This was mainly due to their inability to respond dynamically or use flexible language.

In contrast, AI-powered chatbots based on LLM architectures, such as ChatGPT, have brought a shift in educational interaction by offering contextual adaptability, open-ended dialogue, and generative language capabilities [15], [22], [42]. These systems rely on deep learning, natural language processing, and reinforcement learning from human feedback to simulate conversations that resemble human dialogue across a range of educational settings [13], [21], [24]. Tools like ChatGPT and more advanced versions such as GPT-4 can serve in multiple instructional capacities, including tutoring, summarizing content, generating feedback, supporting academic writing, and simulating clinical scenarios [14], [20], [28].

One of the key features of ChatGPT-style chatbots is their capacity for multi-turn, context-sensitive dialogue. This ability enables them to act as intelligent learning partners instead of simply delivering isolated pieces of information [44], [48]. Unlike rule-based systems, these models dynamically generate responses informed by large-scale training corpora, enabling nuanced interpretation of learner inputs and personalized scaffolding [18], [29]. The multimodal capabilities of newer LLMs, which incorporate text, speech, and even visual data, further enhance their educational potential [11], [13].

Across the corpus of reviews, interaction types enabled by LLMs are notably varied and educationally rich. ChatGPT supports dialogic exchanges, intelligent tutoring systems, collaborative writing, Q&A formats, translation, summarization, simulation of clinical or professional scenarios, and content generation [31], [33], [49]. These functionalities reflect a substantial departure from the constrained dialogues of rule-based chatbots and illustrate the versatility of LLMs in adapting to learners' cognitive, linguistic, and affective needs [35], [39].

Some reviews classify chatbots functionally as task-oriented (e.g., for quiz delivery or tutoring) or non-task-oriented (e.g., open-ended conversational companions), with LLM-based bots increasingly straddling both categories through their flexible use cases [40]. Additionally, studies identify a growing prevalence of chatbots deployed as virtual learning assistants, intelligent feedback agents, and co-creators of educational content, expanding their role in facilitating self-regulated learning and learner autonomy [19], [30]. Figure 5 provides a comparative overview of functional differences between rule-based and LLM-enabled chatbots across educational use cases.

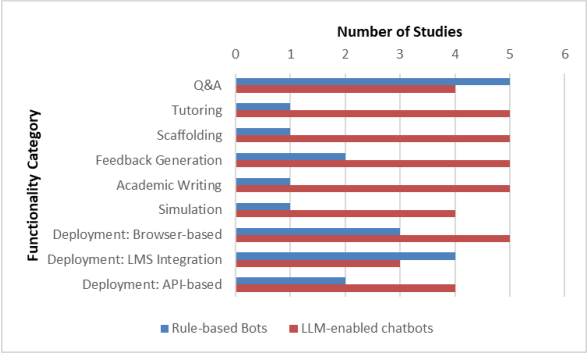

Fig. 5 Functional comparison of rule-based and LLM-enabled chatbots

In terms of deployment, ChatGPT-style chatbots are typically accessed through web-based platforms, mobile apps, and LMS integrations, with APIs and browser interfaces enabling seamless use across diverse digital environments [12], [34]. The platform-agnostic nature of these tools allows them to be embedded in both synchronous and asynchronous learning contexts, thereby enhancing their reach and functionality [41], [44].

A recurring pattern across the literature is the recognition of ChatGPT's transformative potential in education, particularly in higher education and specialized domains such as healthcare, law, and language learning [39], [44], [46]. However, concerns also emerge regarding the accuracy, transparency, and domain specificity of generative outputs, especially in fields requiring high factual precision [34], [46].

Notably, while LLM-based chatbots dominate recent discourse, relatively few reviews systematically compare their pedagogical effectiveness with traditional bots or across different student populations and learning settings. Furthermore, the literature reveals a gap in empirical studies evaluating the longitudinal impact of LLM-enabled chatbot use on learning outcomes, student engagement, and educator roles [20], [45].

D. Learning Subjects and Knowledge Domains

Across the reviewed literature, ChatGPT and related AI chatbot technologies have been deployed in a broad array of academic domains, reflecting their adaptability and potential to

support diverse curricular objectives. However, this versatility is not evenly distributed, as patterns of use vary notably by discipline. A critical synthesis of findings reveals three major clusters of disciplinary engagement: STEM and technical domains, health and medical sciences, and language education. Within and across these clusters, the integration of chatbots, particularly LLM-based systems such as ChatGPT, demonstrates both promising affordances and emergent limitations.

A substantial portion of the literature highlights the integration of chatbots in STEM education, particularly in computer science, mathematics, physics, engineering, and related fields [5], [19], [38], [48]. These subjects benefit from the structured nature of their content, which aligns well with the rule-based and responsive architecture of chatbot interactions. ChatGPT has been used to assist with problem-solving, concept explanation, and programming tasks [42], [23], [26], with students leveraging its capabilities for debugging code, generating solutions, and preparing for STEM assessments [12], [29]. However, while the chatbot's application in quantitative domains is widespread, findings also indicate variability in performance. For example, ChatGPT is reported to perform less effectively in mathematics and software testing than in conceptual fields like economics or critical thinking [16], underscoring limitations in domains requiring high levels of precision and procedural accuracy.

The most consistently documented domain for LLM-based chatbot use is medical and health sciences. Reviews report applications across a wide range of specialties, including nursing, neurosurgery, dentistry, internal medicine, and parasitology [27], [39], [44]. ChatGPT is employed for clinical case simulation, diagnostic reasoning, patient communication, and evidence-based practice [14], [20], [22]. These uses exploit the tool's generative capacity to create realistic scenarios and support clinical decision-making processes. Nonetheless, performance across subdomains can vary significantly, as shown in studies where accuracy differed by anatomical topic [46]. Overall, the emphasis on medical education reflects a clear alignment between chatbot functionalities and the pedagogical needs of content-heavy, procedurally complex fields. At the same time, the focus on healthcare underscores a research gap in other vocational and professional disciplines.

Language education, particularly in English as a Foreign or Second Language (EFL/ESL), represents another key area of chatbot application. ChatGPT and other AI tools have supported learners in writing, speaking, vocabulary development, grammar practice, and pronunciation tasks [14], [32], [45]. The dialogic and feedback-oriented design of chatbots aligns well with communicative language teaching, offering opportunities for personalized and repetitive practice. Moreover, generative capabilities are harnessed for storytelling, dialogue simulation, and cultural literacy tasks [21], [40]. Despite this, reviews consistently note imbalances in coverage, with listening skills remaining underexplored [14] and a lack of studies addressing the full breadth of the humanities. A few reviews report limited chatbot applications in history, literature, philosophy, or ethics [2], [24], pointing to a broader underutilization in discursive and interpretive disciplines.

Several studies emphasize ChatGPT's cross-disciplinary adaptability, often employed in support functions such as academic writing, research design, and curriculum development regardless of specific subject matter [11], [32], [50]. In such cases, ChatGPT is positioned as a general-purpose educational support tool, enhancing skills like argument construction, idea organization, and content synthesis [28], [31]. Some reviews document its use in professional education contexts such as law, journalism, and tourism, albeit with fewer empirical studies [14], [41], [43]. The absence of deeper curricular integration in these domains highlights a research gap in assessing the pedagogical alignment of generative AI with subject-specific learning outcomes.

Notably, while the breadth of disciplinary coverage is wide, the distribution is uneven. STEM and health sciences dominate, whereas social sciences and humanities are relatively underrepresented. This skew may reflect the ease with which chatbot functionalities align with structured and procedural knowledge but also raises concerns about neglecting interpretive, critical, and reflective domains that are central to liberal education. Moreover, many studies report ChatGPT's use in academic writing across domains without sufficiently interrogating how writing conventions differ by discipline [34], [49]. This lack of granularity suggests a need for more nuanced research into how LLM-based tools interact with the epistemological foundations and discursive norms of specific subjects. A summary of the distribution of studies across educational subdomains is provided in Figure 6.

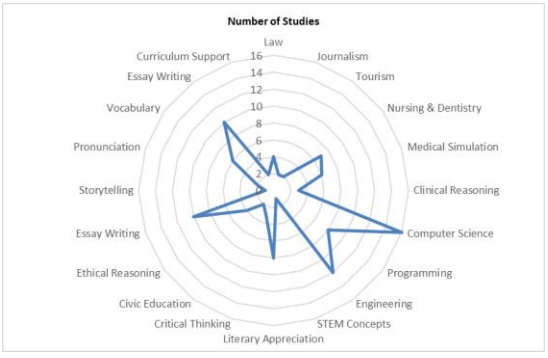

Fig. 6 Number of studies per educational subdomain utilizing AI chatbots

E. The Role of the Educator

The integration of ChatGPT and AI chatbots in education has catalyzed a paradigmatic shift in the role of educators, repositioning them from content transmitters to facilitators, designers, and ethical overseers of human-AI learning environments. This transformation is consistently observed

across the systematic reviews analyzed, though its depth and manifestation vary depending on institutional context, educator preparedness, and the specific chatbot technologies in use.

As ChatGPT and other generative AI tools become capable of producing instructional content, answering student questions, and simulating learning scenarios, educators face the challenge of integrating these technologies into coherent pedagogical frameworks [16], [22], [23]. This new responsibility includes helping students engage critically with AI-generated material, creating assessments that are resistant to automation, and ensuring that human oversight remains central to the learning process [14], [20], [24]. In this evolving context, teachers take on the dual role of designing learning experiences mediated by chatbots and interpreting the insights produced by AI systems [15], [42].

Recent literature highlights the importance of structured professional development that helps educators build skills in prompt engineering, critically assess AI-generated content, and use chatbot data effectively to inform instructional choices [11], [33], [43]. Several review studies recommend training initiatives that go beyond technical proficiency and address the pedagogical and ethical aspects of working with AI tools [5], [28], [29]. Such professional learning is increasingly necessary, especially since teachers may either make limited use of ChatGPT or rely on it excessively when clear pedagogical frameworks are lacking [34], [47].

The role of the educator also involves acting as an ethical guide and protecting academic integrity in learning environments where AI tools are present. Teachers are expected to demonstrate responsible use of ChatGPT, help students understand its limitations and potential biases, and implement measures to prevent plagiarism and misuse [14], [18]. This responsibility also includes rethinking how assessment is designed, ensuring it takes into account the presence of generative AI and encourages students to produce original work and engage in critical thinking [44], [49]. Reviews also highlight the need for educators to interpret AI outputs with a critical lens, ensuring that learners are not misled by incorrect or superficial responses generated by ChatGPT [32], [46].

Educators act as gatekeepers who determine when, how, and to what extent ChatGPT is incorporated into instruction [12], [45]. Their role includes curating content, configuring chatbot interactions, interpreting usage analytics, and scaffolding student engagement [26], [40]. Even in disciplines where ChatGPT demonstrates significant utility, such as medical education or academic writing, educators are critical in validating outputs and contextualizing them within domain-specific epistemologies [27], [39]. The literature suggests that effective integration is most successful when educators maintain an active presence, using chatbots to enhance rather than replace human interaction [30], [36].

Despite widespread consensus on the educator's evolving role, several tensions persist. Many studies report ambivalence or resistance among educators, often stemming from concerns about workload, diminished educator–student interaction, or uncertainty about pedagogical value [2], [17], [48]. While chatbots are frequently portrayed as tools that can reduce cognitive load by automating feedback or routine instruction, their use also raises fears about instructional deskilling and the erosion of pedagogical authority [38], [50]. Furthermore, the literature identifies a lack of empirical research on how educators adapt their practices over time and how different educational cultures or institutional structures shape these adaptations. Few studies systematically examine the long-term impact of professional development interventions or provide comparative insights across educational levels or disciplines. Figure 7 summarizes the key dimensions of educators' evolving roles in AI chatbot-enhanced learning environments, as identified across the reviewed literature.

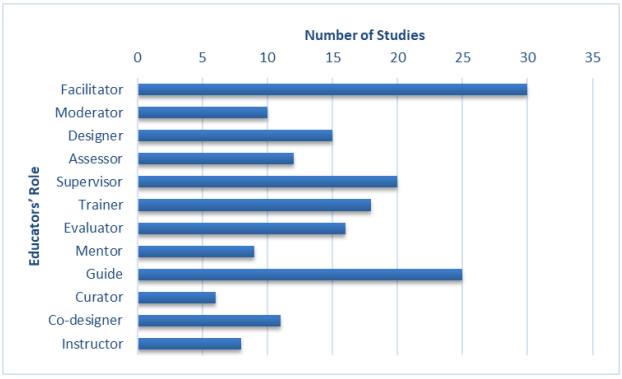

Fig. 7 Educators' evolving roles in AI chatbot-enhanced learning environments

F. Perceptions of Students and Educators

Across the reviewed literature, student and educator perceptions of chatbots in education are characterized by a complex interplay of enthusiasm, caution, and evolving practices. A consistent finding is the broadly positive reception among students, who frequently highlight ChatGPT's immediacy, availability, and personalized feedback as key benefits for learning support, particularly in writing, revision, brainstorming, and language development [11], [14], [16], [18]. These tools are seen as helpful companions in asynchronous and self-paced learning, with students reporting increased motivation, confidence, and autonomy in navigating academic tasks [31], [36], [40]. Such perceptions are shaped not only by the chatbot's functionality but also by learners' digital literacy, language proficiency, and the pedagogical context in which these tools are used [5], [11], [38].

At the same time, concerns around overreliance, reduced cognitive engagement, and the superficial nature of AI-generated responses recur throughout the literature [15], [35], [41]. Students appreciate the convenience and non-judgmental space offered by ChatGPT but also express skepticism about accuracy, originality, and ethical boundaries [23], [28]. These mixed perceptions reflect an increasing awareness among students that using LLMs for academic purposes still requires critical thinking and careful fact-checking [24], [46].

Many educators acknowledge that AI tools can save time and support instruction by offering feedback, responding to common questions, and helping to guide student learning, especially in large classrooms or when resources are limited [2], [48]. However, these benefits are often tempered by concerns over academic integrity, reduced student effort, and challenges to traditional pedagogical authority [15], [17], [39]. Several reviews point to educators' skepticism about the reliability and curricular alignment of chatbot outputs, particularly when used for assessment or critical thinking tasks [14], [26], [41].

Notably, a divide emerges between student and educator attitudes. Students often approach chatbots as pragmatic learning aids, whereas educators are more likely to perceive them through the lens of risk management and pedagogical integrity [22], [50]. This divergence underscores the need for institutional support structures, such as professional development, clear ethical guidelines, and co-constructed usage frameworks that promote transparency and pedagogical alignment [20], [29], [37].

Students are developing adaptive prompting strategies, using ChatGPT for feedback, self-testing, and peer collaboration simulations [44], [45]. Educators, though more cautious, are beginning to adopt blended approaches that combine AI tools with human facilitation to support engagement and personalization [16], [43]. However, most reviews highlight limited evidence on long-term effects, especially concerning learning quality, critical thinking, and socio-emotional outcomes [23], [42]. Student perspectives dominate the discourse, while educator voices remain underrepresented across disciplines and levels [12], [40]. Insights from parents, support staff, and diverse cultural contexts are also scarce. Additionally, many studies fail to clearly distinguish between general chatbots and LLM-based tools like ChatGPT, despite their distinct capabilities and educational implications [19], [32]. A comparative summary of student and educator perceptions is illustrated in Figure 8.

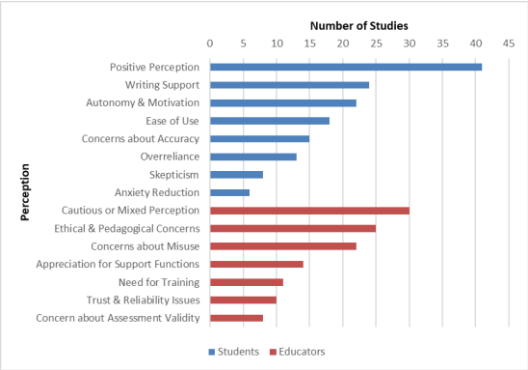

Fig. 8 Student and educator perceptions of AI chatbots in educational contexts

G. Learning Outcomes

The integration of chatbots in education, and particularly ChatGPT and similar LLM systems, has been associated with a broad spectrum of learning outcomes across cognitive, metacognitive, and affective domains. Across the literature, both chatbot-agnostic reviews and ChatGPT-specific studies converge on several key findings, while also revealing important divergences and underexplored areas.

One of the most consistently reported outcomes across reviews is the improvement in academic performance, especially in task-based and content-heavy disciplines such as medicine, STEM, and academic writing [11], [26], [38], [40]. ChatGPT, in particular, has demonstrated efficacy in enhancing students' comprehension, content organization, and written expression [14], [20], [41]. It aids knowledge acquisition through structured feedback, scaffolded learning tasks, and the simplification of complex concepts [14], [22], [30]. Moreover, it facilitates exam preparation and improves task efficiency, with some studies reporting performance levels that rival or surpass those of human learners in specific domains such as medical licensing [27], [33]. Nevertheless, the strength of these claims is tempered by methodological limitations. Many studies rely on subjective self-reports or observational data, with relatively few deploying rigorous experimental or longitudinal designs [12], [33].

ChatGPT has been widely praised for improving academic writing, language fluency, and rhetorical organization; [16], [23], [28]. Its real-time feedback and rephrasing suggestions promote writing fluency and reduce students' cognitive load during composition, which in turn boosts motivation and time-on-task [17], [32]. In ESL/EFL contexts, chatbots were found to yield comparable language development outcomes to human interlocutors, particularly in vocabulary acquisition and speaking fluency when supplemented with transcription tools [32]. Still, over-reliance poses risks. Several reviews caution that ChatGPT may inadvertently suppress students' original thinking and critical engagement with content, especially in writing tasks where AI-generated text might be accepted uncritically [21], [34]. Moreover, while writing assistance is frequently reported, robust evidence on long-term language development remains limited [14].

Across the corpus, there is recurring emphasis on the potential of chatbots to support 21st-century competencies such as critical thinking, creativity, self-regulation, and digital literacy [5], [15], [45]. ChatGPT has been used to encourage reflective practices, analyze ethical dilemmas, and explore counterarguments, thereby fostering higher-order thinking [24], [44]. In simulated environments and writing support tasks, it enables iterative learning and metacognitive engagement [22], [49].

However, the depth of these outcomes varies significantly. In some cases, chatbot use resulted in shallow engagement or rote use of AI responses without critical evaluation [39], [46]. The presence or absence of pedagogical scaffolding appears to be a key determinant: when instructors provide structured prompts and reflection tasks, critical thinking outcomes improve; in unmoderated settings, cognitive dependency is more likely [15], [32].

Both general chatbot systems and ChatGPT specifically

support self-regulated learning by providing timely, individualized feedback and 24/7 access to resources [48], [50]. These affordances enhance learner autonomy, planning, and monitoring, which are core elements of metacognitive development [14], [18]. Reviews also note increased motivation and engagement, particularly when chatbots incorporate gamified elements or simulate social presence [25], [36]. Yet, the benefits are not uniform. Some studies report declines in learner initiative when students overly rely on AI assistance, a pattern that may counteract the intended metacognitive benefits [34], [41].

The overall picture of learning outcomes associated with chatbot use is cautiously optimistic, though not without caveats. While cognitive and affective gains are broadly supported, inconsistencies remain regarding long-term learning, higher-order cognitive processes, and subject-specific efficacy. Particularly for ChatGPT, results are often task-dependent: performance is stronger in structured writing and reasoning tasks than in complex scientific or diagnostic scenarios [42], [46]. Few studies examine the longitudinal effects of chatbot usage on learning trajectories. There is also limited comparative work on how different models (e.g., GPT-3.5 vs. GPT-4.0) influence outcomes across disciplines [27]. A consolidated overview of reported learning outcomes is presented in Figure 9.

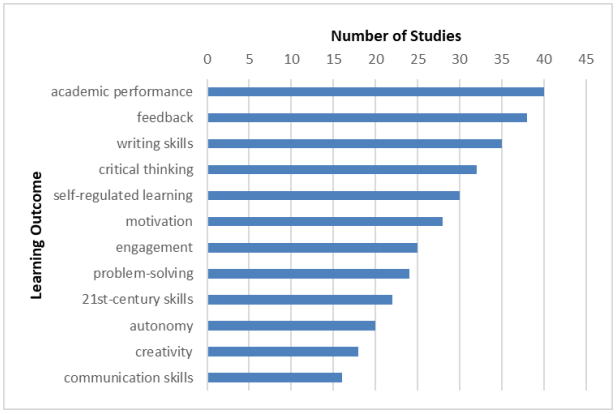

Fig. 9 Reported learning outcomes from the use of AI chatbots in education

H. Benefits and Advantages

A prominent convergence across the reviewed literature is the consistent recognition of the pedagogical, cognitive, and operational benefits associated with AI chatbots in education, particularly with ChatGPT and similar LLM-based systems. These benefits span personalized learning, cognitive scaffolding, instructional efficiency, and learner autonomy, underscoring the transformative potential of chatbot technologies in both general and domain-specific educational contexts.

Several reviews highlight the ability of chatbots to personalize learning by adjusting content, feedback, and pacing based on each learner's profile [12], [22], [45]. ChatGPT, in particular, has been recognized for producing context-sensitive responses that align with different levels of prior knowledge, learning preferences, and language abilities [23], [24], [29]. This kind of personalization supports differentiated instruction and benefits a wide range of learners, including those with special educational needs or limited language proficiency [31], [34]. The ability of large language models to act as tutors, ethical dialogue partners, or collaborative learning companions strengthens their contribution to learner-centered approaches [24].

The continuous availability of chatbots is frequently mentioned as a key benefit, giving learners the opportunity to receive assistance outside of regular classroom hours [36], [38], [50]. This constant access supports learner autonomy by enabling self-directed study, quick resolution of questions, and repeated practice as needed [14], [48]. Particularly in asynchronous or remote contexts, ChatGPT facilitates flexible engagement, reducing dependence on instructor presence and allowing students to direct their own educational trajectories [11], [20], [21].

LLM-based systems support learners in cognitive tasks such as summarization, paraphrasing, argument construction, and idea generation [14], [18], [28]. ChatGPT enhances metacognitive engagement by assisting with planning, organizing, and monitoring academic work, thereby fostering self-regulated learning [42], [49]. These capabilities are particularly valuable in language learning and writing-intensive tasks, where students benefit from immediate, low-stakes environments that reduce anxiety and support repeated practice [25], [32].

Educators benefit from the automation of routine tasks such as grading, content generation, syllabus design, and feedback provision [15], [16], [33]. This reallocation of educator workload allows for greater emphasis on pedagogical design and mentoring [32], [43]. The reviews consistently highlight ChatGPT's role in streamlining academic operations, supporting formative assessment, and offering scalable solutions for large classes [17], [47]. Additionally, chatbot integration has shown promise in administrative functions such as inquiry handling and academic advising [34].

Several studies underscore the role of chatbots in promoting inclusivity, particularly by assisting learners who face linguistic, cognitive, or socio-economic challenges [5], [44]. By facilitating multilingual communication, offering assistive feedback, and enhancing access to educational content, ChatGPT contributes to broader educational equity [31], [39]. Furthermore, its capacity to engage learners through simulated decision-making, conversational scaffolding, and gamified feedback mechanisms has been associated with improved motivation and reduced learner anxiety [19], [40].

The scalability of chatbots enables their effective use across disciplines and educational levels, from language learning and general education to medical training and professional development [27], [46]. ChatGPT's ability to generate complex content such as clinical scenarios, ethical dilemmas, and discipline-specific questions further extends its

applicability [20], [39]. However, several reviews caution that these benefits are contingent on responsible implementation, robust ethical governance, and alignment with pedagogical goals [32], [34]. Figure 10 illustrates the main pedagogical, cognitive, and operational benefits of AI chatbots identified across the reviewed studies.

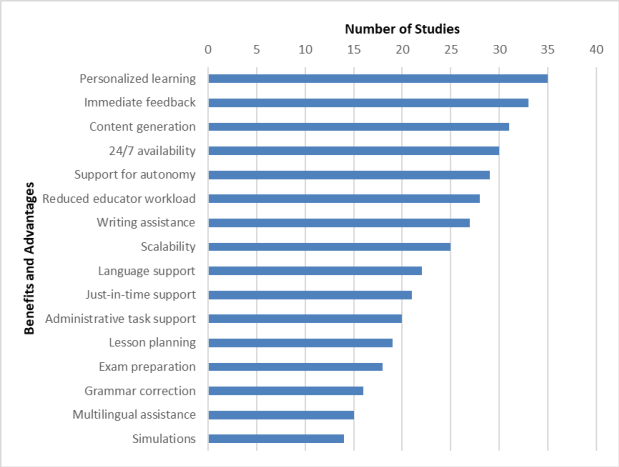

Fig. 10 Educational advantages of AI chatbots

I. *Challenges and Limitations*

The integration of ChatGPT and similar AI chatbots in education is accompanied by a broad range of challenges that span technical, pedagogical, ethical, and contextual dimensions. Across the 41 systematic reviews and meta-analyses synthesized, several converging themes emerge, particularly regarding the reliability of AI-generated content, infrastructural limitations, and the complex sociocultural dynamics that shape chatbot adoption and use.

A prominent and recurrent concern is the phenomenon of hallucinations, where AI systems such as ChatGPT generate factually incorrect, fabricated, or misleading content that nonetheless appears coherent and authoritative. This issue was identified as particularly problematic in knowledge-intensive domains such as healthcare, science, and programming, where factual accuracy is critical [11], [14], [46]. The limitations of ChatGPT's knowledge cutoff, often cited as pre-2022, exacerbate the generation of outdated information [16], [28]. Additionally, citation errors and the absence of source transparency further undermine trust in AI outputs [34], [41].

Pedagogically, overreliance on AI tools emerged as a persistent concern, with multiple studies cautioning that sustained use of chatbots may reduce critical thinking, creativity, and learner autonomy [12], [18], [23]. These risks are compounded by ethical challenges, including plagiarism, manipulation of assessment tasks, and erosion of academic integrity [41], [43], [49]. The inability of current plagiarism detection tools to identify AI-generated content further complicates enforcement efforts [16].

Equity-related challenges were also widely reported. Reviews noted that disparities in digital infrastructure, internet connectivity, and device availability can limit access to AI tools, particularly in under-resourced or rural contexts [5], [20], [39]. Cultural and linguistic biases embedded in training data may marginalize learners from diverse backgrounds, leading to exclusion or misrepresentation [15], [21], [45]. Moreover, language processing limitations hinder the chatbot's performance in multilingual or non-standard linguistic settings [37], [38].

Technological challenges, both in terms of software and hardware, further constrain chatbot utility. These include issues such as poor platform integration, limited interoperability with learning management systems, and inconsistent user experiences due to prompt sensitivity or response variability [29], [30], [47]. Some reviews also highlighted ChatGPT's lack of interactivity and emotional nuance, which can negatively affect learner engagement and hinder socially complex tasks [32], [44].

A commonly reported sociocultural barrier to adoption involved resistance from both educators and students, which stemmed from limited familiarity with the technology, doubts about its educational usefulness, and fears of replacement or professional deskilling [36], [47], [48]. A recurring observation was that educator training on AI literacy is often insufficient or absent, limiting effective and ethical implementation [23], [37].

Across studies, a critical meta-level limitation was noted in the current state of evidence itself. Several reviews pointed to the superficiality of existing empirical work, the predominance of short-term interventions, and a lack of robust evaluation metrics to assess long-term educational impact [14], [33]. Moreover, many studies demonstrated rapid publication cycles with limited peer review, raising questions about the methodological rigor underlying claims of efficacy [32], [33]. Figure 11 summarizes the key pedagogical, technical, ethical, and infrastructural limitations identified across the reviewed studies.

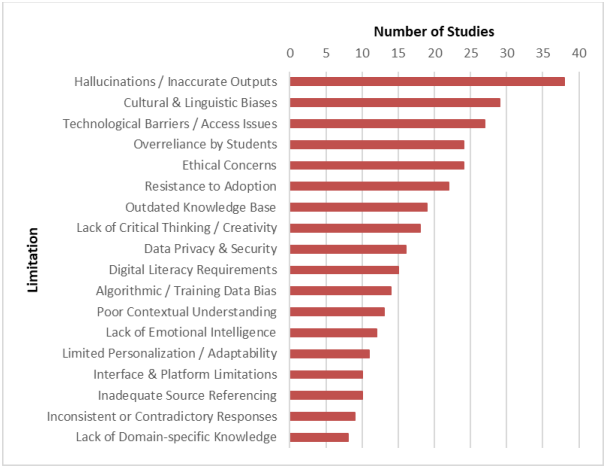

Fig. 11 Key challenges and limitations of AI chatbots in education

J. *Ethical and Legal Issues*

A strong consensus emerges across the reviewed literature regarding the ethical and legal challenges posed by the integration of chatbots in education, particularly with LLMs

like ChatGPT. These concerns cluster into four major thematic domains: academic integrity and authorship, data privacy and consent, transparency and explainability, and institutional preparedness and policy frameworks.

Academic dishonesty, especially in the form of plagiarism, ghostwriting, and inappropriate student reliance on chatbot-generated outputs, represents one of the most frequently cited concerns. Numerous reviews [11], [15], [16], [23], [41] detail how ChatGPT can facilitate unauthorized assistance in assessments and generate content that bypasses standard plagiarism detection tools. The issue is compounded by the tool's inability to cite verifiable sources or differentiate between factual and hallucinated content [21], [46]. Furthermore, ChatGPT has appeared as a co-author in some publications, a practice that has provoked debate over authorship attribution and the ethical boundaries of AI-assisted academic production [28], [44]. These findings underscore the urgent need for academic institutions to clarify the status of AI-generated content in scholarship and to update honor codes and authorship conventions accordingly [42], [49].

Closely linked to ethical deployment is the safeguarding of personal and sensitive data. Reviews emphasize that many chatbot platforms, particularly those using LLMs, operate with limited transparency about how user data is collected, stored, and processed [36], [45], [48]. The risks include breaches of confidentiality, unauthorized data use, and noncompliance with regulations such as the GDPR [2], [32], [43]. Studies further caution that students and educators are often unaware of the privacy implications of chatbot use, especially when AI systems adaptively gather behavioral data for feedback or personalization [37], [47]. Despite the seriousness of these risks, relatively few studies offer in-depth analyses of legal compliance or institutional safeguards, indicating a critical gap in the literature [40].

The black-box nature of generative AI presents another persistent ethical concern. Several reviews emphasize that the opaque nature of LLM algorithms poses a major challenge to explainability, making it difficult for users to understand, verify, or question the outputs generated by AI tools [14], [20], [33]. In the absence of interpretability mechanisms, both educators and learners may be exposed to misinformation, biased content, or conclusions that lack a clear rationale [27], [39]. To address these risks, many studies call for the adoption of explainable AI (XAI) frameworks and user-centered design approaches that promote transparency in how these systems operate and make decisions [19], [43]. The importance of maintaining human oversight and judgment in AI-facilitated learning is also repeatedly stressed, particularly to mitigate the risks of false authority perception and erosion of critical thinking [24], [32].

Many reviews criticize the lack of institutional preparedness to manage the ethical and legal challenges of AI adoption in education. There is a recurring call for comprehensive policy frameworks that govern acceptable use, attribution norms, data protection, and the role of AI in assessment and instruction [12], [17], [26]. Ethical risks are exacerbated by low levels of AI literacy among both educators and students, making it difficult to distinguish between appropriate and inappropriate uses of chatbots [18], [29]. Recommendations include integrating digital ethics into curricula, providing professional development on AI pedagogy, and conducting privacy impact assessments to anticipate and mitigate harms [31], [32]. Notably, several reviews stress the importance of cross-sector collaboration between educators, policymakers, and developers to ensure that regulatory responses are not only reactive but also anticipatory and equity-oriented [34], [50]. Figure 12 provides a thematic overview of the principal ethical and legal concerns associated with AI chatbot use in education, as discussed across the reviewed studies.

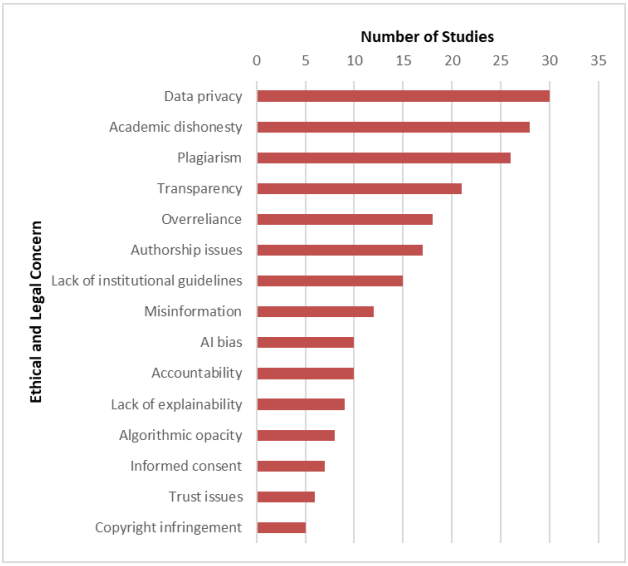

Fig. 12 Key ethical and legal issues in AI chatbots use in education

K. Educational Policy and Institutional Integration

A consistent thread across the reviewed literature is the growing consensus that the integration of AI chatbots, particularly ChatGPT and similar LLMs, demands the establishment of comprehensive educational policy frameworks. These frameworks are seen as essential for governing ethical use, ensuring academic integrity, and aligning AI deployment with curricular and pedagogical goals. Despite the limited explicit referencing of international organizations such as UNESCO or the OECD, the majority of studies converge on the urgency of institutional and national policy responses to guide responsible AI adoption.

A dominant pattern in the literature is the emphasis on institutional-level policymaking as the primary locus of governance for AI integration. Numerous studies call for the development of context-sensitive, transparent, and evidence-based guidelines to regulate the use of ChatGPT in both higher education and K-12 contexts [12], [15], [16], [28], [48]. These policies are expected to address a broad array of concerns, including data privacy, plagiarism, content validity, and equitable access to AI tools [18], [21] [45]. Several reviews

stress the importance of incorporating ethical standards, professional development programs, and AI literacy into institutional policy structures [14], [20], [42].

Across both general chatbot literature and ChatGPT-specific analyses, there is broad agreement on the need for policies that codify ethical principles and pedagogical appropriateness. These include developing clear rules for AI-assisted assessments, setting expectations around academic honesty, and clarifying the limits of chatbot contributions to student work [23], [32], [49]. A key concern is the balance between automation and human interaction, with several studies cautioning against the replacement of educators with AI tools. Instead, policies are urged to promote complementary integration, leveraging chatbots for repetitive or low-stakes tasks while preserving human-led instruction for more nuanced, interactive learning [24], [32].

While few studies explicitly engage with UNESCO or OECD policy frameworks, many imply their relevance by highlighting the absence of standardized international guidelines [34], [40], [50]. This gap is especially noted in the context of cross-border digital education strategies and ethical governance, where alignment with global educational goals such as SDG4 is considered beneficial [31], [43]. Several reviews indirectly advocate for global benchmarking, calling for policies that ensure fairness, transparency, and adaptability across diverse educational systems [5], [32].

Another recurrent theme is the call for participatory approaches to policy development. Reviews advocate for the co-design of policies involving educators, students, administrators, and technologists to ensure that frameworks are grounded in pedagogical realities and attuned to emerging AI capabilities [17], [36], [41]. This collaborative model is often framed as necessary to navigate the fast-evolving landscape of generative AI, where institutional reflexivity and ongoing evaluation are indispensable [39], [46], [47].

Several reviews highlight a critical implementation gap: the lack of educator readiness for AI integration. Accordingly, they recommend that policy frameworks include structured professional development and AI literacy initiatives for educators and academic leaders [11], [19], [22]. Without sufficient training, even the most well-crafted policies risk poor adoption or misuse. In addition, tailored guidelines are needed for domain-specific contexts, such as nursing, health sciences, and medical education, where AI raises unique ethical and procedural concerns [20], [27], [39].

Despite general agreement on the necessity of policy development, several contradictions emerge. For instance, while some studies call for swift policy enactment to match the pace of AI innovation [14], [44], others advocate for a more cautious, evidence-informed approach that delays formal integration until further research validates pedagogical outcomes [46]. Moreover, very few reviews engage with the role of chatbot technologies in non-tertiary or resource-constrained settings, marking a critical underexplored area [30], [38]. Figure 13 highlights the institutional policy dimensions identified as critical for the ethical, pedagogically aligned, and scalable integration of AI chatbots in education.

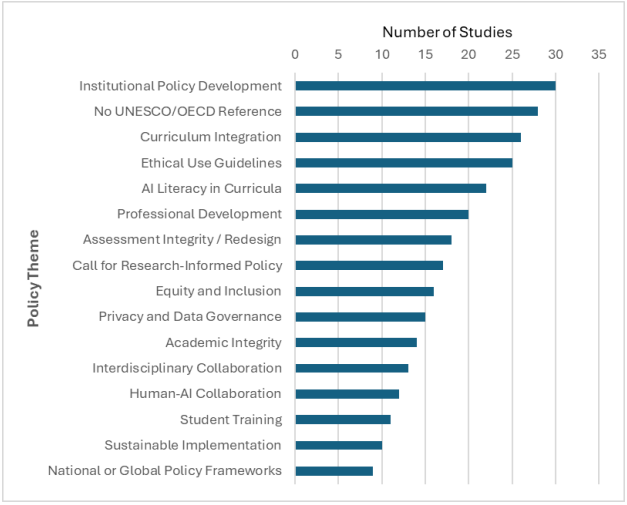

Fig. 13 Institutional policy priorities for the responsible integration of AI chatbots in education

IV. DISCUSSION

The principal objective of this umbrella review was to synthesize and critically appraise extant systematic reviews, meta-analyses, and meta-syntheses concerning the educational applications of ChatGPT and related AI-driven chatbots. Despite the rapid proliferation of such technologies in educational discourse and practice, prior reviews have largely been fragmented, often limited in disciplinary scope, geographically concentrated in high-income Anglophone contexts, and lacking longitudinal or cross-level analyses. Furthermore, much of the existing literature has focused on exploratory studies with limited methodological rigor, leaving critical gaps in understanding the pedagogical efficacy, ethical implications, and policy readiness associated with chatbot integration across diverse educational settings. This umbrella review addresses these gaps by providing an integrative and thematically structured synthesis of 41 peer-reviewed studies, offering a comprehensive account of current trends, limitations, and future directions for research, practice, and policymaking in AI-supported education.

A. Pedagogical Functions and Disciplinary Contexts of LLM-Based Chatbots

In response to RQ1, "What types and pedagogical functions of chatbots are identified in the literature, and in which educational contexts and subject domains are they applied?", the findings demonstrate that AI chatbots, particularly ChatGPT, are predominantly deployed within tertiary education, often embedded in digitally enriched learning environments. This predominance reflects not only the infrastructural and curricular affordances of higher education institutions but also, as [51] suggest the discursive positioning of universities as innovation labs for educational technology. Such environments foster iterative adoption cycles and

legitimize experimental pedagogies, yet this framing privileges techno-optimistic narratives and potentially sidelines critical scrutiny of equity and access. As [52] argue, institutional contexts mediate students' engagement with AI in ways that are deeply entangled with cultural, socioeconomic, and cognitive variables, suggesting that "facilitating conditions" are not neutral enablers but socially patterned constraints. Furthermore, [53] highlights that the habitual use of AI tools and their perceived usefulness are often driven not by deliberate pedagogical strategies, but by factors such as convenience, gamified interaction, and external pressures related to academic performance. This raises important concerns about the instrumental use of such technologies in education. Additionally, the analysis by [54] challenges assumptions about the easy transferability of chatbot practices, demonstrating that significant differences in institutional cultures and priorities make cross-contextual comparisons difficult to sustain.

The use of chatbots in education appears to be most prominent in disciplines such as health sciences, STEM subjects, and language learning. In medical education, ChatGPT is commonly employed to support clinical reasoning and examination preparation, while in STEM disciplines, it aids in programming, logical reasoning, and problem-solving tasks. Language learning contexts benefit from the chatbot's capacity for vocabulary expansion, writing assistance, and conversational fluency, functions that capitalize on its generative and dialogic capabilities. These patterns reflect broader educational logics in which automation augments practices of structured reasoning, personalized scaffolding, and dialogic rehearsal. In STEM and computing contexts, [55] conceptualize the chatbot not merely as a technical support tool but as a cognitive partner that mediates learners' engagement with algorithmic structures and abstract problem-solving. This arrangement shifts pedagogical agency toward AI-assisted learning environments. Similarly, [56] highlights how language education repurposes chatbot interaction as a site of affective engagement and authentic communication, reconfiguring communicative competence around sustained, low-stakes conversational practice with artificial interlocutors. In medical training, [57] foreground the chatbot's capacity to simulate diagnostic reasoning as a pedagogical device, but also raise critical concerns about the epistemic status of AI-generated responses and the implications for clinical judgment.

In contrast, humanities and social sciences remain relatively underexplored domains, despite offering fertile ground for critical discourse, ethical deliberation, and reflective thinking. This underutilization of AI in the humanities and social sciences runs counter to interpretive perspectives that view such technologies not merely as tools for automation, but as agents that shape epistemic practices and pedagogical relations. [58] argue that while AI can replicate formal structure and coherence in academic writing, it lacks the critical reflexivity and situated understanding essential to humanistic inquiry, thereby raising concerns about the erosion of authentic assessment and the narrowing of intellectual depth. [59], writing from a sociological perspective, questions the assumption that AI systems are emotionally neutral. He argues that the design of these tools often reflects subtle yet significant assumptions about authority, agreement, and emotional tone in the construction of knowledge. These issues become particularly evident in disciplines such as law or philosophy. In these fields, as [60] point out, AI-supported learning may help students perform better in tasks without necessarily fostering deeper cognitive autonomy or ethical reasoning.

In addition, the literature makes a clear distinction between rule-based chatbot systems and those powered by large language models. Rule-based systems are effective for delivering predefined, predictable interactions, but they offer limited flexibility. In contrast, LLM-driven chatbots such as ChatGPT allow for open-ended conversations, respond to context, and can engage with users through multiple modes of communication. These capabilities represent a shift away from simple automation and toward the augmentation of learning, supporting more learner-centered and inquiry-based educational approaches. This shift to LLM-driven chatbots aligns with pedagogical debates on learner agency and reframes teaching as a dialogic, co-constructive process. [61] frame such AI integration not merely as a tool for efficiency but as a catalyst for reimagining autonomy and cognitive engagement within educational ecosystems. However, this optimism is tempered by critical perspectives that view LLMs as both enabling and constraining, as [62] suggest that while these systems scaffold complex task performance, they also risk over-structuring learning through algorithmic patterning. Meanwhile, [63] expose tensions between pedagogical aspirations and institutional realities, cautioning that the dialogic potential of ChatGPT is entangled with issues of authorship, assessment validity, and digital inequity. Together, these perspectives illuminate not only the opportunities but also the contested terrains of policy, pedagogy, and power reconfigured by generative AI.

B. Learning Outcomes Across Cognitive, Affective, and Behavioral Domains

Addressing RQ2, "What kinds of learning outcomes are associated with the use of chatbots in education, including cognitive, affective, and behavioral dimensions?", the evidence reviewed indicates that chatbot use is generally associated with improved learning outcomes, particularly in structured and task-oriented domains such as medicine and engineering. ChatGPT appears to enhance academic writing and reading comprehension by scaffolding learner cognition and providing iterative, individualized feedback. Yet, as several studies suggest, these learning benefits cannot be interpreted in isolation from the pedagogical and institutional frameworks within which chatbot use unfolds. [64] emphasize that without methodological structures that frame the chatbot as a means of cultivating creativity and critical thinking, its use risks reinforcing superficial learning. Nugroho et al. [65]

similarly argue that while cognitive gains are evident, they become educationally meaningful only when learners are guided to engage with the tool reflexively, confronting issues of credibility, authorship, and academic responsibility. Echoing this, [66] underline the need for instructional models that reposition the educator not as a transmitter of knowledge, but as a mediator of autonomous, AI-supported learning.

Moreover, several reviews identify gains in metacognitive awareness, learner autonomy, and digital literacy. The availability of 24/7 support, personalized feedback, and just-in-time assistance contributes to the development of self-regulated learning. However, the degree of metacognitive engagement varies considerably across contexts. In the absence of educator mediation, some learners may develop an overreliance on chatbot-generated outputs, potentially compromising critical thinking. This concern echoes those raised by [67], who warned that using ChatGPT as a self-directed learning tool, without pedagogical guidance, may result in superficial understanding and undetected misconceptions.

Additionally, the effectiveness of ChatGPT is highly task-dependent. It performs reliably in activities that involve structured reasoning and language production, but demonstrates notable limitations in mathematics and content areas requiring factual precision. The current body of evidence also lacks longitudinal data, making it difficult to assess sustained learning gains, especially in relation to critical thinking and socio-emotional development. According to [68], addressing this gap is essential for understanding the evolving nature of ChatGPT adoption, as his longitudinal research demonstrates that students' engagement with the tool significantly changes over time, influenced by shifting levels of trust, emotional responses, and perceived behavioral control.

C. Perceptions and Pedagogical Roles in Chatbot-Supported Learning

In response to RQ3, "How do students and educators perceive the benefits and limitations of chatbots, and in what ways is the role of the educator evolving in AI-supported learning environments?", the reviewed studies indicate that students generally report favorable attitudes toward ChatGPT. Cited advantages include immediacy, anonymity, non-judgmental feedback, and increased confidence in performing academic tasks. Students frequently use the tool for drafting assignments, revising texts, and preparing for assessments. Importantly, many students also express awareness of ChatGPT's limitations, such as concerns about accuracy, ethical usage, and the potential for dependency. This emergent form of AI literacy reflects a broader shift toward critical engagement, wherein learners do not merely acquire operational competence but begin to interrogate the epistemic and ethical dimensions of their interactions with generative tools. [69], approaching the topic from the perspective of ethics in health education, emphasizes the importance of integrating transparency, informed consent, and data ownership into the educational use of AI. He cautions that allowing these systems to become normalized without oversight could undermine student autonomy. [70] also highlight a broader institutional hesitation. While many educators recognize the pedagogical potential of AI, they often lack the necessary preparation to manage its ethical integration. This points to a worrying gap between the rapid spread of AI technologies and the slower, less coordinated pace of curricular reform. [61] propose a systems-level view, suggesting that AI literacy should be seen not merely as a technical ability but as a critical skill that enables users to identify the limitations, biases, and opaque operations of LLMs. At the same time, [71] raise questions about the concept of student empowerment, noting that frequent use of AI tools often coincides with a lack of critical reflection.

Educators tend to approach the use of generative AI with greater caution and critical reflection. Although many recognize its potential to ease their workload and improve instructional efficiency, they also express serious concerns about academic integrity, the diminishing role of teacher authority, and the reliability of AI-generated content. A significant number of educators feel unprepared to incorporate chatbot technologies into their teaching in meaningful ways, which underscores the urgent need for specialized training and professional development. As [72] points out, the gap between students' enthusiasm and teachers' hesitation is not just a matter of differing adoption speeds. It reflects deeper shifts in how authority, accountability, and educational purpose are being reshaped in the context of AI-driven learning.

D. Ethical Considerations and Institutional Policy Responses to Educational Chatbots

In response to RQ4, "What ethical concerns and institutional or policy-level responses are discussed in relation to the integration of chatbots into formal education?", the literature highlights academic integrity as a central issue, with plagiarism and ghostwriting emerging as primary concerns. The difficulty ChatGPT has in reliably citing sources or verifying the accuracy of its claims further complicates its role in high-stakes assessments. Privacy risks are also frequently mentioned, particularly in relation to the lack of transparency around data storage and user profiling by providers of large language models. This lack of transparency not only limits the explainability of these systems but also undermines ethical standards in educational accountability. [73] argues that this situation points to a need for AI literacy that equips both educators and students with the skills for epistemic transparency and moral reasoning. [74] similarly describes the opaque nature of these systems as a structural vulnerability that disrupts institutional understandings of authorship, responsibility, and academic integrity in contexts shaped by algorithmic collaboration.

Few institutions have established clear policies governing acceptable use, assessment protocols, or educator responsibilities in AI-enhanced learning. Moreover, significant equity gaps persist. Access to AI tools remains uneven,

particularly in low-resource settings and among linguistically marginalized populations. While several reviews advocate for ethical guidelines and educator training, only a minority engage with international standards or propose adaptive governance mechanisms responsive to local needs. This policy vacuum exacerbates existing digital divides by reinforcing systemic inequalities in who can access, benefit from, and influence the use of AI technologies, an issue [75] frames as a failure to anticipate ethical responsibilities in education, [76] highlight as privileging already dominant voices in academic and technological spaces, and [77] interpret as a lack of inclusive, context-sensitive governance mechanisms needed to ensure equitable innovation**.**

V. IMPLICATIONS

The findings of this umbrella review yield several theoretical, practical, and policy-level implications. Theoretically, the widespread integration of ChatGPT in higher education and professional training reinforces constructivist models of learning, where dialogic engagement and personalized scaffolding play central roles. However, the uneven adoption across disciplines and educational levels highlights a gap in current learning theories that must better account for AI-mediated instruction in underrepresented settings, including primary and special education.

Practically, educators should be encouraged to adopt ChatGPT as a supplementary pedagogical tool that enhances learner autonomy, feedback immediacy, and metacognitive engagement. To implement these tools effectively, educators need access to well-structured professional development programs. Such programs should go beyond teaching the technical aspects of AI and prompt engineering, placing equal emphasis on developing a critical understanding of the ethical and epistemological limits of these technologies. Teachers must be prepared to address challenges related to overdependence on AI, the spread of misinformation, and the diminishing role of student authorship in learning processes.

From a policy perspective, institutions should develop flexible governance frameworks that address academic integrity, data privacy, transparency, and equitable access. These frameworks need to be proactive and inclusive, involving educators, students, technologists, and policymakers in their design. Because generative AI has global implications, institutional strategies can benefit from alignment with established international guidelines such as UNESCO's Recommendation on the Ethics of Artificial Intelligence [8], the EU Artificial Intelligence Act [9], and the OECD AI Principles [10]. At the same time, the integration of AI chatbots into education brings important ethical concerns related to authorship, fairness, and learner autonomy. Addressing these concerns requires clear guidelines and open, inclusive dialogue among stakeholders. Without such efforts, there is a risk of reinforcing existing inequalities or weakening pedagogical approaches that promote critical thinking.

VI. LIMITATIONS

This umbrella review is subject to several limitations. One key issue is that most of the included studies come from high-income, English-speaking countries, which may limit how well the findings apply to regions with different cultural and educational contexts. Another limitation concerns the nature of the original studies, many of which relied on short-term, exploratory research designs. These often used self-reported data and did not include longitudinal approaches, making it difficult to draw conclusions about long-term impacts. Third, there is considerable variability in the definitions, scopes, and evaluation criteria employed across the source studies, complicating direct comparison and synthesis. Additionally, the rapid pace of LLM development means that newer tools and use cases may not yet be captured in the evidence base. To mitigate these limitations, we employed a comprehensive search strategy across multiple databases and applied rigorous inclusion criteria to select high-quality reviews. However, the inherent limitations of the source studies constrain the conclusions that can be drawn.

Future research should address the methodological and contextual gaps identified in this review. Specifically, studies are needed from a broader range of geographical and cultural settings, including underrepresented and non-Anglophone regions. There is also a pressing need for longitudinal research designs capable of capturing sustained learning outcomes over time, as well as greater standardization in definitions, inclusion criteria, and evaluation metrics. These limitations constrain the generalizability and comparability of current findings, particularly across diverse educational systems. Addressing them will enhance the robustness and practical applicability of future research in the evolving field of AI-supported education.

VII. CONCLUSION

This umbrella review synthesized findings from 41 systematic reviews, scoping reviews, and meta-analyses on the integration of ChatGPT and related LLM-based chatbots in educational settings. The analysis offers a comprehensive overview of their pedagogical potential, methodological limitations, and socio-technical implications. Overall, the evidence portrays a dual landscape: on one side, there is considerable enthusiasm for the educational affordances of chatbots, particularly in enhancing feedback, engagement, and personalized learning. On the other side, there are significant concerns related to ethics, equity, and the uneven maturity of research and policy frameworks.

The findings converge on several key pedagogical benefits, particularly in higher education and professional domains such as STEM, health, and language learning. Chatbots are shown to support cognitive development, writing proficiency, and learner autonomy through scalable, dialogic, and metacognitively rich interactions. These technologies have demonstrated value as adaptive learning companions in both synchronous and asynchronous environments.

Although there is a sense of optimism surrounding the educational use of LLMs, several important limitations need to be acknowledged. Many of the reviewed studies rely on exploratory or short-term research designs, offering limited evidence about long-term impacts or cross-contextual comparisons. The fast pace of LLM development has outstripped efforts to validate their effectiveness, leaving critical questions unanswered regarding their lasting influence on learning outcomes, disciplinary understanding, and critical thinking. In addition, empirical research on primary education, special education, and underrepresented regions remains scarce, raising concerns about equity and the broader applicability of current findings.

The role of educators is widely recognized as crucial, yet it often remains underexplored. While LLM-based chatbots are reshaping instruction toward more interactive and student-centered models, many teachers feel unprepared to use them effectively. The literature stresses the importance of professional development that builds both technical and ethical understanding. At the same time, student and teacher perceptions reveal tensions in AI adoption. Students describe benefits like increased confidence and motivation but also express concerns about overdependence and unreliable content. Educators, though cautiously open, worry about academic integrity and maintaining coherence in teaching. These dynamics underscore the need for collaborative and transparent approaches that align AI use with sound pedagogical principles.

Despite the rapid uptake of generative AI in education, significant ethical and infrastructural concerns persist. Issues such as system opacity, data privacy, authorship, and unequal access continue to raise questions about fairness and accountability. These challenges are often exacerbated by the lack of cohesive policy responses, both at the institutional and international levels. Although many studies call for ethical guidelines and collaboration among stakeholders, few actively engage with established global frameworks like those of UNESCO or the OECD. As a result, existing governance efforts remain fragmented and reactive. What is urgently needed is a more coordinated and participatory approach to policymaking, one that is informed by empirical evidence and attuned to the practical realities of teaching and learning.

These findings need to be viewed in light of important limitations, such as the overrepresentation of studies from high-income, English-speaking contexts and the lack of consistency in research methods across the literature. Moving forward, it is essential that future research not only addresses these methodological and geographical gaps, but also examines broader questions related to ethical governance, the changing role of educators in AI-supported classrooms, and the wider social and political dimensions of integrating AI into education systems. A more interdisciplinary, equity-oriented, and policy-aware research agenda will be essential. Ultimately, understanding how generative chatbots reshape teaching and learning requires not only empirical vigilance but also normative reflection on the future of education in the age of AI.

Declaration of Generative AI and AI-assisted Technologies in the Writing Process

During the preparation of this work the authors used ChatGPT4o in order to assist with refining content. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

References

[1] Abd-Alrazaq et al., "Large language models in medical education: opportunities, challenges, and future directions," JMIR Med. Educ., vol. 9, no. 1, p. e48291, 2023, doi: 10.2196/48291.

[2] H. Lee, "The rise of ChatGPT: Exploring its potential in medical education," Anat. Sci. Educ., vol. 17, no. 5, pp. 926–931, 2024, doi: 10.1002/ase.2270.

[3] D. R. Cotton, P. A. Cotton, and J. R. Shipway, "Chatting and cheating: Ensuring academic integrity in the era of ChatGPT," Innov. Educ. Teach. Int., vol. 61, no. 2, pp. 228–239, 2024, doi: 10.1080/14703297.2023.2190148.

[4] B. Eager and R. Brunton, "Prompting higher education towards AI-augmented teaching and learning practice," J. Univ. Teach. Learn. Pract., vol. 20, no. 5, pp. 1–19, 2023, doi: 10.53761/1.20.5.02.

[5] M. Dimeli and A. Kostas, "The Role of ChatGPT in Education: Applications, Challenges: Insights From a Systematic Review," J. Inf. Technol. Educ. Res., vol. 24, p. 2, 2025, doi: 10.28945/5422.

[6] Y. K. Dwivedi et al., "Opinion Paper: 'So what if ChatGPT wrote it?' Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy," Int. J. Inf. Manage., vol. 71, p. 102642, 2023, doi: 10.1016/j.ijinfomgt.2023.102642.

[7] K. B. Ooi et al., "The potential of generative artificial intelligence across disciplines: Perspectives and future directions," J. Comput. Inf. Syst., vol. 65, no. 1, pp. 76–107, 2025, doi: 10.1080/08874417.2023.2261010.

[8] UNESCO, "Recommendation on the ethics of artificial intelligence," United Nations Educational, Scientific and Cultural Organization, 2021. [Online]. Available: https://unesdoc.unesco.org/ark:/48223/pf0000381137/PDF/381137eng.pdf.multi. Accessed: Jul. 17, 2025.

[9] European Parliament and Council of the European Union, "Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending certain Union legislative acts (Artificial Intelligence Act)," Off. J. Eur. Union, L 2024/1689, 2024. [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=OJ:L_202401689. Accessed: Jul. 17, 2025.

[10] OECD, "Recommendation of the Council on Artificial Intelligence," OECD Legal Instruments, 2019. [Online].

Available: https://oecd.ai/en/assets/files/OECD-LEGAL-0449-en.pdf. Accessed: Jul. 17, 2025.

[11] S. Dube, et al., "Students' perceptions of ChatGPT in education: A rapid systematic literature review," in Intelligent Computing. SAI 2024, K. Arai, Ed., Lecture Notes in Networks and Systems, vol. 1019, Cham, Switzerland: Springer, 2024, pp. 258–279, doi: 10.1007/978-3-031-62273-1_18.

[12] M. I. Baig and E. Yadegaridehkordi, "ChatGPT in the higher education: A systematic literature review and research challenges," Int. J. Educ. Res., vol. 127, p. 102411, 2024, doi: 10.1016/j.ijer.2024.102411.

[13] C. K. Lo, P. L. H. Yu, S. Xu, D. T. K. Ng, and M. S. Y. Jong, "Exploring the application of ChatGPT in ESL/EFL education and related research issues: a systematic review of empirical studies," Smart Learn. Environ., vol. 11, no. 1, p. 50, 2024, doi: 10.1186/s40561-024-00342-5.

[14] X. Xu, Y. Chen, and J. Miao, "Opportunities, challenges, and future directions of large language models, including ChatGPT in medical education: a systematic scoping review," J. Educ. Eval. Health Prof., vol. 21, pp. 1–11, 2024, doi: 10.3352/jeehp.2024.21.6.

[15] A. N. Ansari, S. Ahmad, and S. M. Bhutta, "Mapping the global evidence around the use of ChatGPT in higher education: A systematic scoping review," Educ. Inf. Technol., vol. 29, no. 9, pp. 11281–11321, 2024, doi: 10.1007/s10639-023-12223-4.

[16] C. K. Lo, "What is the impact of ChatGPT on education? A rapid review of the literature," Educ. Sci., vol. 13, no. 4, p. 410, 2023, doi: 10.3390/educsci13040410.

[17] B. Amarathunga, "ChatGPT in education: unveiling frontiers and future directions through systematic literature review and bibliometric analysis," Asian Educ. Dev. Stud., vol. 13, no. 5, pp. 412–431, 2024, doi: 10.1108/AEDS-05-2024-0101.

[18] W. S. Yun and S. Surianshah, "A review of the scholarly works on ChatGPT use in education: Bibliometric analysis," Int. J. Technol. Educ., vol. 7, no. 3, pp. 650–666, 2024, doi: 10.46328/ijte.823.

[19] D. F. Al Husaeni, N. Haristiani, W. Wahyudin, and R. Rasim, "Chatbot artificial intelligence as educational tools in science and engineering education: A literature review and bibliometric mapping analysis with its advantages and disadvantages," ASEAN J. Sci. Eng., vol. 4, no. 1, pp. 93–118, 2024, doi: 10.17509/ajse.v4i1.67429.

[20] J. Gunawan, Y. Aungsuroch, and J. Montayre, "ChatGPT integration within nursing education and its implications for nursing students: A systematic review and text network analysis," Nurse Educ. Today, vol. 141, p. 106323, 2024, doi: 10.1016/j.nedt.2024.106323.

[21] T. W. Kim, "Application of artificial intelligence chatbots, including ChatGPT, in education, scholarly work, programming, and content generation and its prospects: a narrative review," J. Educ. Eval. Health Prof., vol. 20, p. 38, 2023, doi: 10.3352/jeehp.2023.20.38.

[22] A. A. Abujaber, A. Abd-Alrazaq, A. R. Al-Qudimat, A. J. Nashwan, and A. AbuJaber, "A strengths, weaknesses, opportunities, and threats (SWOT) analysis of ChatGPT integration in nursing education: a narrative review," Cureus, vol. 15, no. 11, p. e48643, 2023, doi: 10.7759/cureus.48643.

[23] M. Montenegro-Rueda, J. Fernández-Cerero, J. M. Fernández-Batanero, and E. López-Meneses, "Impact of the implementation of ChatGPT in education: A systematic review," Computers, vol. 12, no. 8, p. 153, 2023, doi: 10.3390/computers12080153.

[24] M. Mahsun, M. Ali, I. R. Ekaningrum, and H. Ibda, "Trend of using ChatGPT in learning process and character education: A systematic literature review," Int. J. Learn. Teach. Educ. Res., vol. 23, no. 5, pp. 387–402, 2024, doi: 10.26803/ijlter.23.5.20.

[25] B. Li, V. L. Lowell, C. Wang, and X. Li, "A systematic review of the first year of publications on ChatGPT and language education: Examining research on ChatGPT's use in language learning and teaching," Comput. Educ.: Artif. Intell., vol. 7, p. 100266, 2024, doi: 10.1016/j.caeai.2024.100266.

[26] X. Deng and Z. Yu, "A meta-analysis and systematic review of the effect of chatbot technology use in sustainable education," Sustainability, vol. 15, no. 4, p. 2940, 2023, doi: 10.3390/su15042940.

[27] E. G. Özgür and N. Bekiroğlu, "Can ChatGPT be used as an education assistant in the field of health sciences? Examining with the integration of meta-analysis and co-citation analysis," Heliyon, vol. 11, no. 3, 2025, doi: 10.1016/j.heliyon.2025.e41925.

[28] M. Imran and N. Almusharraf, "Analyzing the role of ChatGPT as a writing assistant at higher education level: A systematic review of the literature," Contemp. Educ. Technol., vol. 15, no. 4, p. ep464, 2023, doi: 10.30935/cedtech/13605.

[29] P. Zhang and G. Tur, "A systematic review of ChatGPT use in K- 12 education," Eur. J. Educ., vol. 59, no. 2, p. e12599, 2024, doi: 10.1111/ejed.12599.

[30] C. K. Nee, M. H. A. Rahman, N. Yahaya, N. H. Ibrahim, R. A. Razak, and C. Sugino, "Exploring the trend and potential distribution of chatbot in education: A systematic review," Int. J. Inf. Educ. Technol., vol. 13, no. 3, pp. 516–525, 2023, doi: 10.18178/ijiet.2023.13.3.1834.

[31] E. Faisal, "Unlock the potential for Saudi Arabian higher education: a systematic review of the benefits of ChatGPT," in Front. Educ., vol. 9, p. 1325601, Apr. 2024, doi: 10.3389/feduc.2024.1325601.

[32] J. Jeon and S. Lee, "Can learners benefit from chatbots instead of humans? A systematic review of human-chatbot comparison research in language education," Educ. Inf. Technol., vol. 29, pp. 23329–23360, 2024, doi: 10.1007/s10639-024-12725-9.

[33] A. Aster, M. C. Laupichler, T. Rockwell-Kollmann, G. Masala, E. Bala, and T. Raupach, "ChatGPT and Other Large Language Models in Medical Education—Scoping Literature Review," Med. Sci. Educ., vol. 35, pp. 555–567, 2024, doi: 10.1007/s40670-024-02206-6.

[34] M. Li, A. Enkhtur, F. Cheng, and B. A. Yamamoto, "Ethical implications of ChatGPT in higher education: A

scoping review," J. Interdiscip. Stud. Educ., vol. 13, no. 1, pp. 55–68, 2024, doi: 10.32674/jise.v13i1.6072.

[35] I. M. García-López, C. S. G. González, M. S. Ramírez-Montoya, and J. M. Molina-Espinosa, "Challenges of implementing ChatGPT on education: Systematic literature review," Int. J. Educ. Res. Open, vol. 8, p. 100401, 2025, doi: 10.1016/j.ijedro.2024.100401.

[36] Y. Albadarin, M. Saqr, N. Pope, and M. Tukiainen, "A systematic literature review of empirical research on ChatGPT in education," Discover Educ., vol. 3, no. 1, p. 60, 2024, doi: 10.1007/s44217-024-00138-2.

[37] D. Ramandanis and S. Xinogalos, "Designing a chatbot for contemporary education: A systematic literature review," Information, vol. 14, no. 9, p. 503, 2023, doi: 10.3390/info14090503.

[38] I. Maita, S. Saide, A. M. Putri, and D. Muwardi, "ProsCons of Artificial Intelligence-ChatGPT Adoption in Education Settings: A Literature Review and Future Research Agendas," IEEE Eng. Manag. Rev., vol. 52, no. 3, pp. 27–42, 2024, doi: 10.1109/EMR.2024.3394540.

[39] M. Zarei, M. Zarei, S. Hamzehzadeh, S. S. B. Oliyaei, and M. S. Hosseini, "ChatGPT, a friend or a foe in medical education: A review of strengths, challenges, and opportunities," Shiraz E-Med. J., vol. 25, no. 6, p. e145840, 2024, doi: 10.5812/semj-145840.

[40] O. Chamorro-Atalaya et al., "Application of the chatbot in university education: A systematic review on the acceptance and impact on learning," Int. J. Learn. Teach. Educ. Res., vol. 22, no. 9, pp. 156–178, 2023, doi: 10.26803/ijlter.22.9.9.

[41] A. R. Vargas-Murillo, I. N. M. de la Asuncion, and F. de Jesús Guevara-Soto, "Challenges and opportunities of AI-assisted learning: A systematic literature review on the impact of ChatGPT usage in higher education," Int. J. Learn. Teach. Educ. Res., vol. 22, no. 7, pp. 122–135, 2023, doi: 10.26803/ijlter.22.7.7.

[42] H. K. Park and S. N. Martin, "Exploring the Role of ChatGPT in Science Education for Asia-Pacific and Beyond: a Systematic Review," Asia-Pac. Sci. Educ., vol. 1(aop), pp. 1–32, 2024, doi: 10.1163/23641177-bja10079.

[43] A. A. Fenta, "A review on enhancing education with AI: exploring the potential of ChatGPT, Bard, and generative AI," Discover Educ., vol. 4, no. 1, p. 38, 2025, doi: 10.1007/s44217-025-00426-5.

[44] S. Shorey, C. Mattar, T. L. B. Pereira, and M. Choolani, "A scoping review of ChatGPT's role in healthcare education and research," Nurse Educ. Today, vol. 135, p. 106121, 2024, doi: 10.1016/j.nedt.2024.106121.

[45] S. Salih et al., "Transforming education with AI: A systematic review of ChatGPT's role in learning, academic practices, and institutional adoption," Results Eng., vol. 25, p. 103837, 2025, doi: 10.1016/j.rineng.2024.103837.

[46] D. Chytas et al., "Can ChatGPT play a significant role in anatomy education? A scoping review," Morphologie, vol. 109, no. 365, p. 100949, 2025, doi: 10.1016/j.morpho.2025.100949.

[47] A. Küçükkaya, E. Arikan, and P. Goktas, "Unlocking ChatGPT's potential and challenges in intensive care nursing education and practice: A systematic review with narrative synthesis," Nurs. Outlook, vol. 72, p. 102287, 2024, doi: 10.1016/j.outlook.2024.102287.

[48] K. Naznin, A. Al Mahmud, M. T. Nguyen, and C. Chua, "ChatGPT integration in higher education for personalized learning, academic writing, and coding tasks: A systematic review," Computers, vol. 14, no. 2, p. 53, 2025, doi: 10.3390/computers14020053.

[49] M. Pradana, H. P. Elisa, and S. Syarifuddin, "Discussing ChatGPT in education: A literature review and bibliometric analysis," Cogent Educ., vol. 10, no. 2, p. 2243134, 2023, doi: 10.1080/2331186X.2023.2243134.

[50] A. Čep and A. Bernik, "ChatGPT and Artificial Intelligence in Higher Education: Literature Review Powered by Artificial Intelligence," in Intelligent Computing. SAI 2024, K. Arai, Ed., Lecture Notes in Networks and Systems, vol. 1018, Cham, Switzerland: Springer, 2024, pp. 240–248, doi: 10.1007/978-3-031-62269-4_17.

[51] C. Y. Lai, K. Y. Cheung, and C. S. Chan, "Exploring the role of intrinsic motivation in ChatGPT adoption to support active learning: An extension of the technology acceptance model," Comput. Educ.: Artif. Intell., vol. 5, p. 100178, 2023, doi: 10.1016/j.caeai.2023.100178.

[52] M. Abdaljaleel et al., "A multinational study on the factors influencing university students' attitudes and usage of ChatGPT," Sci. Rep., vol. 14, no. 1, p. 1983, 2024, doi: 10.1038/s41598-024-52549-8.

[53] A. Strzelecki, "To use or not to use ChatGPT in higher education? A study of students' acceptance and use of technology," Interact. Learn. Environ., vol. 32, no. 9, pp. 5142–5155, 2024, doi: 10.1080/10494820.2023.2209881.

[54] A. Strzelecki and S. ElArabawy, "Investigation of the moderation effect of gender and study level on the acceptance and use of generative AI by higher education students: Comparative evidence from Poland and Egypt," Br. J. Educ. Technol., vol. 55, no. 3, pp. 1209–1230, 2024, doi: 10.1111/bjet.13425.

[55] R. Yilmaz and F. G. K. Yilmaz, "The effect of generative artificial intelligence (AI)-based tool use on students' computational thinking skills, programming self-efficacy and motivation," Comput. Educ.: Artif. Intell., vol. 4, p. 100147, 2023, doi: 10.1016/j.caeai.2023.100147.

[56] J. Jeon, "Exploring AI chatbot affordances in the EFL classroom: Young learners' experiences and perspectives," Comput. Assist. Lang. Learn., vol. 37, no. 1–2, pp. 1–26, 2024, doi: 10.1080/09588221.2021.2021241.

[57] R. A. Khan, M. Jawaid, A. R. Khan, and M. Sajjad, "ChatGPT-Reshaping medical education and clinical management," Pak. J. Med. Sci., vol. 39, no. 2, p. 605, 2023, doi: 10.12669/pjms.39.2.7653.

[58] T. Revell et al., "ChatGPT versus human essayists: an exploration of the impact of artificial intelligence for authorship and academic integrity in the humanities," Int.

J. Educ. Integr., vol. 20, no. 1, p. 18, 2024, doi: 10.1007/s40979-024-00161-8.

[59] A. Balmer, "A sociological conversation with ChatGPT about AI ethics, affect and reflexivity," Sociology, vol. 57, no. 5, pp. 1249–1258, 2023, doi: 10.1177/00380385231169676.

[60] S. He and Y. Lu, "The effectiveness of Gen AI in assisting students' knowledge construction in humanities and social sciences courses: learning behaviour analysis," Interact. Learn. Environ., vol. 32, no. 10, pp. 7041–7062, 2024, doi: 10.1080/10494820.2024.2415444.

[61] E. Kasneci et al., "ChatGPT for good? On opportunities and challenges of large language models for education," Learn. Individ. Differ., vol. 103, p. 102274, 2023, doi: 10.1016/j.lindif.2023.102274.

[62] M. M. Rahman and Y. Watanobe, "ChatGPT for education and research: Opportunities, threats, and strategies," Appl. Sci., vol. 13, no. 9, p. 5783, 2023, doi: 10.3390/app13095783.

[63] J. Rudolph, S. Tan, and S. Tan, "War of the chatbots: Bard, Bing Chat, ChatGPT, Ernie and beyond. The new AI gold rush and its impact on higher education," J. Appl. Learn. Teach., vol. 6, no. 1, pp. 364–389, 2023, doi: 10.37074/jalt.2023.6.1.23.

[64] E. Vázquez-Cano, J. M. Ramirez-Hurtado, J. M. Saez-Lopez, and E. Lopez-Meneses, "ChatGPT: The brightest student in the class," Think. Skills Creat., vol. 49, p. 101380, 2023, doi: 10.1016/j.tsc.2023.101380.

[65] A. Nugroho, E. Andriyanti, P. Widodo, and I. Mutiaraningrum, "Students' appraisals post-ChatGPT use: Students' narrative after using ChatGPT for writing," Innov. Educ. Teach. Int., vol. 62, no. 2, pp. 499–511, 2025, doi: 10.1080/14703297.2024.2319184.

[66] W. R. A. Bin-Hady, A. Al-Kadi, A. Hazaea, and J. K. M. Ali, "Exploring the dimensions of ChatGPT in English language learning: A global perspective," Library Hi Tech, 2023, doi: 10.1108/LHT-05-2023-0200.

[67] A. Stojanov, "Learning with ChatGPT 3.5 as a more knowledgeable other: An autoethnographic study," Int. J. Educ. Technol. High. Educ., vol. 20, no. 1, p. 35, 2023, doi: 10.1186/s41239-023-00404-7.

[68] A. Polyportis, "A longitudinal study on artificial intelligence adoption: understanding the drivers of ChatGPT usage behavior change in higher education," Front. Artif. Intell., vol. 6, p. 1324398, 2024, doi: 10.3389/frai.2023.1324398.

[69] K. Masters, "Ethical use of artificial intelligence in health professions education: AMEE Guide No. 158," Med. Teach., vol. 45, no. 6, pp. 574–584, 2023, doi: 10.1080/0142159X.2023.2186203.

[70] D. Lee et al., "The impact of generative AI on higher education learning and teaching: A study of educators' perspectives," Comput. Educ.: Artif. Intell., vol. 6, p. 100221, 2024, doi: 10.1016/j.caeai.2024.100221.

[71] A. Stojanov, Q. Liu, and J. H. L. Koh, "University students' self-reported reliance on ChatGPT for learning: A latent profile analysis," Comput. Educ.: Artif. Intell., vol. 6, p. 100243, 2024, doi: 10.1016/j.caeai.2024.100243.

[72] C. K. Y. Chan, "A comprehensive AI policy education framework for university teaching and learning," Int. J. Educ. Technol. High. Educ., vol. 20, no. 1, p. 38, 2023, doi: 10.1186/s41239-023-00408-3.

[73] T. K. Chiu, "The impact of Generative AI (GenAI) on practices, policies and research direction in education: A case of ChatGPT and Midjourney," Interact. Learn. Environ., vol. 32, no. 10, pp. 6187–6203, 2024, doi: 10.1080/10494820.2023.2253861.

[74] M. Perkins, "Academic Integrity considerations of AI Large Language Models in the post-pandemic era: ChatGPT and beyond," J. Univ. Teach. Learn. Pract., vol. 20, no. 2, pp. 1–24, 2023, doi: 10.53761/1.20.02.07.

[75] C. Kooli, "Chatbots in education and research: A critical examination of ethical implications and solutions," Sustainability, vol. 15, no. 7, p. 5614, 2023, doi: 10.3390/su15075614.

[76] B. D. Lund et al., "ChatGPT and a new academic reality: Artificial Intelligence- written research papers and the ethics of the large language models in scholarly publishing," J. Assoc. Inf. Sci. Technol., vol. 74, no. 5, pp. 570–581, 2023, doi: 10.1002/asi.24750.

[77] B. C. Stahl and D. Eke, "The ethics of ChatGPT–Exploring the ethical issues of an emerging technology," Int. J. Inf. Manage., vol. 74, p. 102700, 2024, doi: 10.1016/j.ijinfomgt.2023.102700.

Contribution of individual authors to the creation of a scientific article (ghostwriting policy)

Emmanouil D. Milakis conducted the data curation and formal analysis, developed the methodology, created visualizations, and was responsible for the original draft and the overall investigation. Constantina Corazon Argyrakou contributed to the validation and provided key resources, while also reviewing and editing the manuscript. Alexandros Melidis participated in data curation and validation, and was involved in the review and editing of the manuscript. John Vrettaros initiated the conceptual framework and methodology, supervised the project, administered the research process, and contributed to the investigation.

Sources of funding for research presented in a scientific article or scientific article itself

No funding was received for conducting this study.

Conflicts of Interest

The authors have no conflicts of interest to declare that are relevant to the content of this article.

APPENDIX

This appendix provides a detailed overview of the methodological features of the 41 review studies analyzed in this umbrella review. Table 1 below presents key characteristics of each included study, including authorship, review type, and number of primary studies analyzed. This supplementary information is intended to enhance transparency, allow for cross-study comparison, and support the interpretive depth of the findings discussed in the main text.

Table I. Methodological Characteristics of the Included Review Studies

Reference No.	Author(s)	Year	Review Type	No. of Studies
[22]	Abujaber et al.	2023	Narrative Literature Review	4600
[19]	Al Husaeni et al.	2024	Bibliometric & Literature Review	376
[36]	Albadarin et al.	2024	Systematic Literature Review	14
[17]	Amarathunga	2024	Bibliometric & Literature Review	45
[15]	Ansari et al.	2024	Systematic Scoping Review	50
[33]	Aster et al.	2024	Scoping Review	145
[12]	Baig & Yadegaridehkordi	2024	Systematic Literature Review	57
[40]	Chamorro-Atalaya et al.	2024	Systematic Literature Review	16
[50]	Čep & Bernik	2024	Systematic Literature Review	Not reported
[46]	Chytas et al.	2025	Scoping Review	9
[26]	Deng & Yu	2023	Meta-analysis	32
[5]	Dimeli & Kostas	2025	Systematic Literature Review	50
[11]	Dube et al.	2024	Rapid Systematic Review	32
[31]	Faisal	2024	Systematic Literature Review	52
[43]	Fenta	2025	Systematic Literature Review	33
[35]	García-López et al.	2025	Systematic Literature Review	42
[20]	Gunawan et al.	2024	Systematic Literature Review	46
[28]	Imran & Almusharraf	2023	Systematic Literature Review	30
[32]	Jeon & Lee	2024	Systematic Narrative Review	11
[21]	Kim	2023	Narrative Literature Review	Not reported
[47]	Küçükkaya et al.	2024	Systematic Literature Review	33
[25]	Li et al.	2024	Systematic Literature Review	36
[34]	Li et al.	2024	Scoping Review	26
[16]	Lo	2023	Rapid Review	50
[13]	Lo et al.	2024	Systematic Literature Review	70
[38]	Maita et al.	2024	Systematic Literature Review	40
[24]	Mahsun et al.	2024	Systematic Literature Review	30
[23]	Montenegro-Rueda et al.	2023	Systematic Literature Review	12
[48]	Naznin et al.	2025	Systematic Literature Review	52
[30]	Nee et al.	2023	Systematic Literature Review	25
[27]	Özgür & Bekiroğlu	2025	Meta-analysis	9
[42]	Park & Martin	2024	Systematic Literature Review	14
[49]	Pradana et al.	2023	Bibliometric Review	93
[37]	Ramandanis & Xinogalos	2023	Systematic Literature Review	73
[45]	Salih et al.	2025	Systematic Literature Review	28
[44]	Shorey et al.	2024	Scoping Review	100
[14]	Xu et al.	2024	Systematic Scoping Review	113
[41]	Vargas-Murillo et al.	2024	Systematic Literature Review	16
[18]	Yun & Surianshah	2024	Bibliometric Review	58
[29]	Zhang & Tur	2024	Systematic Literature Review	13
[39]	Zarei et al.	2024	Narrative Literature Review	Not reported