

Methodological trends and pedagogical applications of ChatGPT in education: An umbrella review of systematic reviews

Abdelghani Babori ^{1*} 

¹The Higher School of Education and Training, Ibnou Zohr University, Agadir, MOROCCO

*Corresponding Author: a.babori@uiz.ac.ma

Citation: Babori, A. (2026). Methodological trends and pedagogical applications of ChatGPT in education: An umbrella review of systematic reviews. *Pedagogical Research*, 11(4), em0291. <https://doi.org/10.29333/pr/19252>

ARTICLE INFO

Received: 18 May 2026

Accepted: 02 Aug. 2026

ABSTRACT

As one of the artificial intelligence tools used worldwide, ChatGPT is increasingly being integrated by users (teachers and learners) into their teaching and/or learning processes. This trend has attracted the attention of education researchers analyzing its benefits and limitations in pedagogical practices. This study presents an umbrella review examining trends in ChatGPT research in education, exploring the bibliographic characteristics of systematic reviews, the methodological approaches employed, and the pedagogical applications of ChatGPT. A systematic search following the PRISMA technique was conducted by querying Scopus, Web of Science, ScienceDirect and ERIC. The twenty-seven reviews selected according to inclusion and exclusion criteria were subjected to in-depth analysis. The results highlight a predominance of systematic reviews, and methodological quality assessment is not systematically conducted. From a pedagogical perspective, ChatGPT is primarily used for scientific writing assistance, pedagogical design support, virtual tutoring, or automated assessment and feedback, providing recommendations for future research.

Keywords: ChatGPT, umbrella review, systematic review, artificial intelligence in education

INTRODUCTION

Artificial Intelligence in Education: Context and Rapid Development

In recent decades, information and communication technologies (ICTs) have evolved rapidly across various sectors of life (health, agriculture, administration, education, etc.). In particular, in the educational domain, these technological innovations impact traditional pedagogical practices while enabling pedagogical innovations (MOOCs, flipped classrooms, hybrid classrooms, distance learning, etc.). However, technological innovation does not always lead to pedagogical innovation. While educational policies view digital tools as a means to modernize teaching, some scientific research highlights a discrepancy between the promises of these technological innovations and their actual use in the classroom (Bernard & Fluckiger, 2019). Artificial intelligence (AI), as a subfield of computer science, has been defined in various ways across the scholarly literature. McCarthy (2007, p. 2) described it as “the science and engineering of making intelligent machines, especially intelligent computer programs.” More broadly, Joyce et al. (2021) characterize AI as an artificial entity capable of meeting or exceeding task requirements while accounting for cultural and demographic contexts. In the institutional domain, the European Parliament characterizes AI as a machine’s ability to display human-like capabilities such as reasoning, learning, planning, and creativity (European Parliament, 2020). Similarly, generative AI is characterized by users’ ability to interact with the system using natural language, as if speaking to a human — an interaction commonly referred to as a “prompt”. Like predictive AI, generative AI draws on training data, learning from a vast number of examples.

ChatGPT in Education: Documented Affordances, Limitations, and the State of Existing Reviews

ChatGPT, developed by OpenAI and launched in November 2022 (Pan et al., 2023), is among the most widely used generative AI applications. Based on the GPT architecture, ChatGPT reportedly reached 100 million users worldwide within two months of its launch (Milmo, 2023). Its ease of use and free access have contributed to its widespread adoption, particularly in educational institutions. An increasing body of research highlights the pedagogical benefits of ChatGPT, notably in essay writing, pedagogical

design, and learning support (Javaid et al., 2023; Jmoula, 2023). Its contribution goes further still. Generative AI can personalize learning pathways to learners' specific needs (Edhuman, 2024), improve accessibility for students with disabilities, and support intelligent tutoring that helps fill learning gaps. It also lightens teachers' cognitive workload by automating the preparation of pedagogical resources and quizzes (de la Higuera & Iyer, 2024). Moreover, several authors have highlighted the contribution of generative AI to pedagogical innovation through the design of immersive learning scenarios and the promotion of inquiry-based active learning, which enables learners to construct knowledge by confronting real-world problem situations (Piekoszewski-Cuq, 2024). However, the literature has consistently emphasized significant limitations and risks associated with the use of ChatGPT in education, concomitantly with its reported affordances. Systematic literature reviews highlight algorithmic biases, hallucinations, and risks related to academic integrity, notably plagiarism and cheating (Atiyah, 2023; Deng et al., 2024; Li et al., 2023).

The research on the use of ChatGPT for educational purposes has experienced particularly rapid growth since late 2022. This surge has produced a growing body of systematic reviews and meta-analyses examining research trends, benefits, and risks associated with its use. These literature reviews have explored a variety of themes, including design and assessment, personalized learning, coding assessment, research and editing, and administrative tasks (Baig & Yadegaridehkordi, 2024). While existing studies identify the primary pedagogical uses of ChatGPT, few propose a comparative analysis of the reviews themselves, particularly regarding their methodological and bibliographic characteristics. Furthermore, the methodological quality of existing literature reviews has rarely been assessed using standardized appraisal tools, thereby complicating the assessment of their findings' reliability.

Purpose of the Study and Research Questions

As primary research and subsequent systematic reviews on ChatGPT in education continue to proliferate, an umbrella review represents the logical next step to address review redundancy and resolve discrepancies across existing syntheses (Conn & Coon Sells, 2014). According to Conn and Coon Sells (2014), such a review provides scientists with directions for future research and offers policy makers evidence-based recommendations. Positioned at the top of the literature review hierarchy, this type of research represents one of the most rigorous and robust syntheses available. A crucial methodological contribution of this approach lies primarily in the rigorous management of overlap between different systematic reviews. Indeed, in the field of artificial intelligence in education, several systematic reviews may examine the same primary studies (cohort studies, expert opinions, etc.) (Abdellatif et al., 2025), which risks biasing conclusions by incorporating duplicate results. More specifically, by going beyond merely summarizing findings, this systematic review enables the identification of publication biases, while drawing on the methodological quality assessment of included reviews using standardized tools such as AMSTAR, AMSTAR 2, or ROBIS.

Beyond synthesizing the existing review literature, this umbrella review makes four complementary contributions to research on ChatGPT in education. First, it provides a second-order synthesis of published reviews rather than primary studies, enabling the identification of convergent findings, methodological inconsistencies, and research gaps across the field. Second, it systematically integrates the AMSTAR 2 appraisal into the interpretation of the evidence, allowing methodological quality to inform the synthesis. Third, it identifies and synthesizes the main pedagogical applications of ChatGPT into five recurring domains across the review literature. Finally, it advances a techno-pedagogical perspective by framing ChatGPT as a component of instructional design rather than an isolated technological tool.

In light of the above, this research proposes an umbrella review that synthesizes and compares existing systematic reviews. More specifically, this research seeks to answer the following questions:

RQ1: What are the bibliographic characteristics of the published reviews?

RQ2: What are the methodological characteristics of these reviews, and how is their quality assessed?

RQ3: What pedagogical uses of ChatGPT are identified in the literature?

RESEARCH METHODOLOGY

The review protocol was retrospectively registered on the Open Science Framework (OSF) prior to manuscript submission. The registration was conducted after data collection in order to enhance transparency and reproducibility of the research process. All methodological materials, including the search strategy, inclusion and exclusion criteria, and data extraction procedures, are available on OSF: https://osf.io/v3qus/overview?view_only=7eaea05b98b24d50aea7eb989b9d5809.

Search Strategy

The research was conducted by consulting the following databases: Scopus, Web of Science, ERIC, and ScienceDirect. These databases were selected because the articles indexed in them are reliable and have undergone a rigorous peer-review process. Some databases, such as ACM and IEEE Xplore, were not considered for two reasons: (a) the majority of their articles are already indexed by Scopus and/or Web of Science, and (b) they focus more on technical research questions related to the use of ChatGPT.

Data collection was carried out between April and June 2025. The keywords used, the specific queries applied in each of these databases, and the number of articles found are described in detail in **Table 1**.

Table 1. Database search queries and results

Database	Query	Results
Scopus	(TITLE-ABS ("ChatGPT") AND TITLE-ABS ("education") AND TITLE-ABS ("systematic review" "scoping review" OR "literature review" OR "umbrella review" OR "meta-analysis")) AND (LANGUAGE (english) OR LANGUAGE (french)) AND (LIMIT-TO (LANGUAGE , "English")) AND (LIMIT-TO (DOCTYPE , "ar") OR LIMIT-TO (DOCTYPE , "re"))	244
Web of science	TS=("ChatGPT") AND TS=("education") AND TS=("systematic review" OR "literature review" OR "scoping review" OR "meta-analysis" OR "umbrella review") AND (LA=("English") OR LA=("French")) AND DT=("Article" OR "Review")	185
ERIC	("ChatGPT" AND "education") AND ("systematic review" OR "literature review" OR "scoping review" OR "meta-analysis" OR "umbrella review") in:title,abstract	60
ScienceDirect	ChatGPT AND education AND (systematic review OR literature review OR scoping review OR meta-analysis OR umbrella review)	89

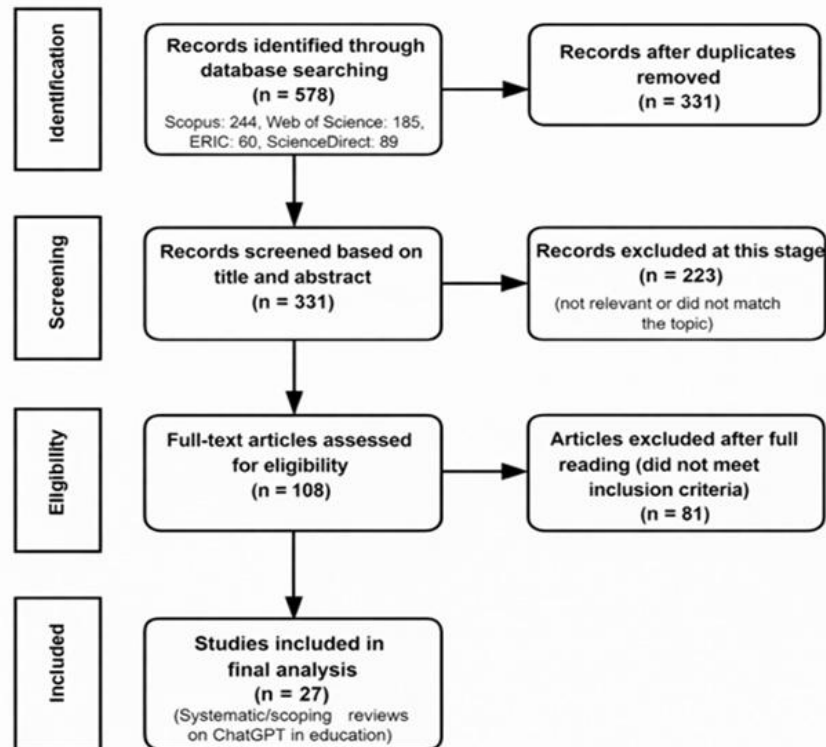
**Figure 1.** Prisma flow diagram of study selection (Adapted from Huang et al., 2025)

Figure 1 presents the main steps of the adopted PRISMA research methodology. The inclusion criteria applied at this first stage (identification) were as follows: (a) article written in English or French; (b) full text availability. No restrictions were imposed regarding the publication date of the articles. In total, after duplicate removal, this search yielded 331 articles, which were retained for further selection in the screening stage.

The second stage aimed to examine the titles and abstracts of the articles identified in the identification phase to exclude irrelevant documents. Articles were retained if they met the following inclusion criteria: (a) correspond to a systematic review or scoping review related to the research questions identified in our study on the pedagogical uses of ChatGPT, (b) the article must be published in a journal: conference proceedings, books, errata, editorials, etc., were excluded. These criteria led to the exclusion of 223 articles during the selection phase. The remaining 108 articles were deemed adequate for a full-text review to assess their eligibility for inclusion in the final phase. During this final phase, an additional criterion was adopted: (c) the review must provide distinct contributions while minimizing significant bibliographic overlap with the other included studies. After thorough examination of the full texts, 81 articles were excluded because they did not meet the previously defined inclusion and exclusion criteria. At the end of this final step, 27 articles were retained for this umbrella review. **Appendix A** lists all 27 reviews included in the corpus in accordance with PRISMA reporting standards; not all are individually cited in the narrative synthesis, as findings are reported at the aggregate/thematic level.

Quality Appraisal of Included Reviews

To ensure the credibility of the data from this umbrella review, AMSTAR 2 was used to assess the methodological quality of the 27 selected systematic reviews. This tool has been widely validated and accepted, and comprises 16 items, including six critical points to consider (Fauvet & Gatignol, 2022): 1) registration of a protocol prior to conducting the review; 2) adequacy of the

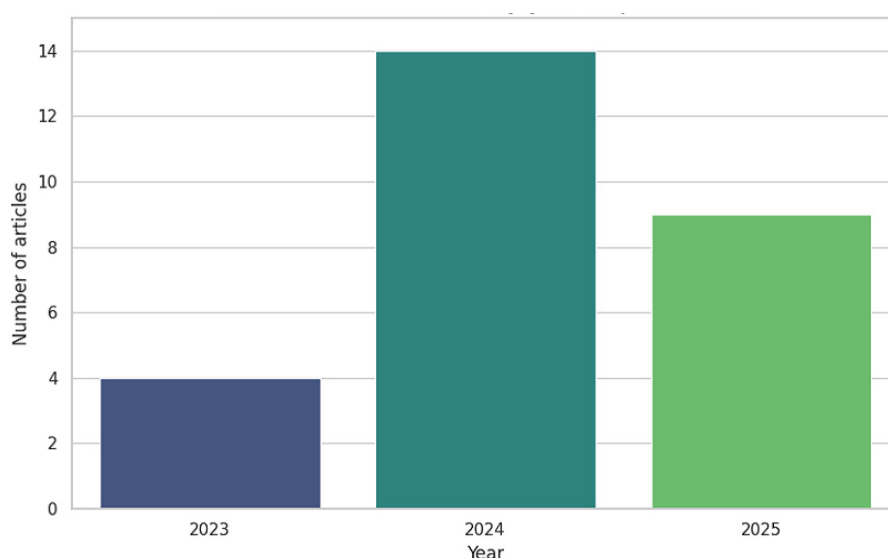


Figure 2. Distribution of articles by year of publication (Source: Author's own elaboration)

bibliographic search; 3) justification for the exclusion of certain studies, 4) assessment of the risk of bias for each included study; 5) selection of an appropriate meta-analysis method; 6) consideration of risks of bias when interpreting the review. Each review was evaluated and classified into one of the following four levels of confidence: high overall confidence, moderate overall confidence, low overall confidence, and critically low overall confidence.

Analysis Method

The synthesis was informed by the AMSTAR 2 quality appraisal. Findings from all eligible reviews were considered; however, greater interpretative weight was given to the five reviews rated as having moderate methodological quality (Bettayeb et al., 2024; Deng et al., 2024; Heung & Chiu, 2025; Fernández Cerero et al., 2025; Park & Martin, 2024), while evidence from lower-rated reviews was used to complement and contextualize the findings.

The twenty-seven selected reviews were subjected to an in-depth analysis using a combined quantitative and qualitative approach. For the quantitative analysis, descriptive statistics were employed to present the distribution of articles by publication year, type of literature review, and number of publications per journal. Subsequently, a qualitative analysis was conducted to describe research trends in terms of research methodologies (data collection and analysis) as well as the pedagogical uses of ChatGPT. Crucially, the findings from the AMSTAR 2¹ appraisal informed the synthesis process. While all twenty-seven reviews were included to provide a comprehensive overview of the subject under study, the results from the five reviews identified as having “moderate overall confidence” were prioritized to ensure the relevance and credibility of the pedagogical recommendations.

A thematic content analysis was applied, inspired by the approach proposed by Hasni et al. (2016). Specifically, excerpts from the corpus were analyzed and segmented into units of meaning, which were then organized into thematic categories describing the various pedagogical uses identified. This analysis also allowed the identification of certain gaps in literature.

RESULTS

Bibliographic Characteristics of The Reviews (RQ1)

To address RQ1, we examined the bibliographic characteristics of the published reviews, focusing on the distribution of articles by year, journal, and geographical location. **Figure 2** details the number of articles published each year from 2023 to 2025. A significant number of articles were published in 2024. From 2025 onwards, this number begins to decline, which may indicate that research in this field is entering a phase of stabilization or maturation.

An analysis of the corpus reveals a clear dominance of systematic reviews ($n = 18$), followed by a smaller subset of scoping reviews ($n = 6$), whereas meta-analyses and bibliometric studies remain notably underrepresented (see **Figure 3**). Despite the emerging nature of ChatGPT research within educational contexts, scoping reviews have managed to maintain a moderate presence. This trend likely stems from a focus on addressing specific, targeted research questions rather than attempting an exhaustive mapping of the entire field. Furthermore, as Arksey and O'Malley (2005) observe, the distinction between various review types is often blurred in practice. They note that the “systematic” label is frequently applied to literature reviews that may lack a

¹ <https://amstar.ca/Amstar-2.php>

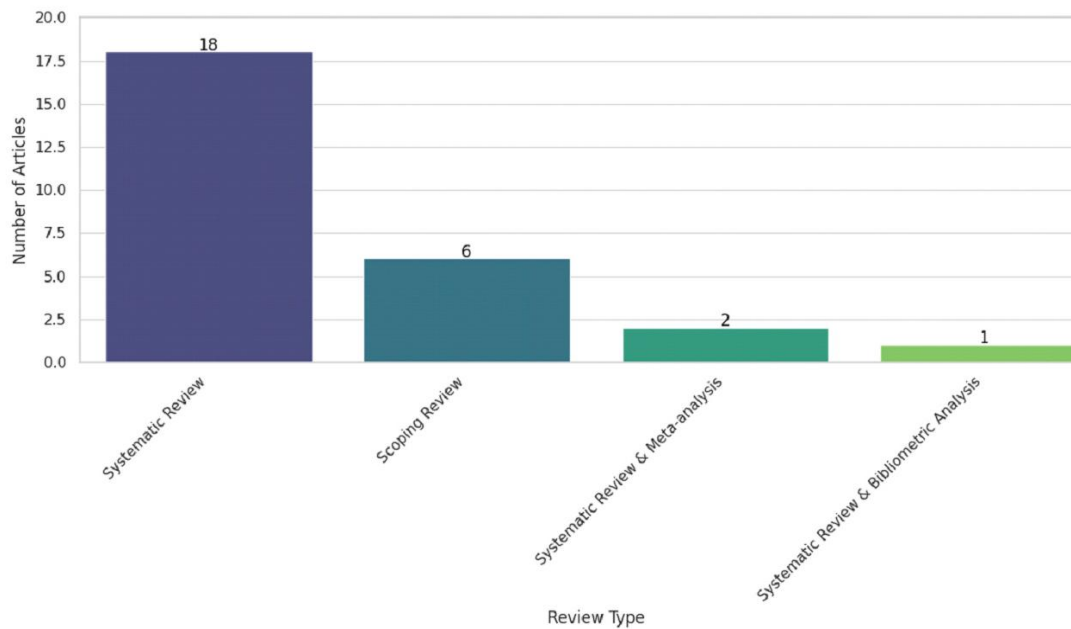


Figure 3. Distribution of articles by review type (Source: Author's own elaboration)

Table 2. Distribution of publications by journal

Journal	Number of publications
Frontiers in Education	2
Computers and Education: Artificial Intelligence	2
Contemporary Educational Technology	2
International Journal of Learning, Teaching and Educational Research	2
Journal of Information Systems Engineering and Business Intelligence	1
TechTrends	1
Education and Information Technologies	1
European Journal of Education	1
Journal of Educational Evaluation for Health Professions	1
Discover Education	1
Nurse Education Today	1
Journal of University Teaching and Learning Practice	1
International Journal of TESOL Studies	1
Computers and Education	1
Assessment & Evaluation in Higher Education	1
Journal of Information Technology Education: Research	1
Journal of Interdisciplinary Studies in Education	1
Technologies	1
Asia-Pacific Science Education	1
Morphologie	1
Communications in Mathematics and Applications	1
South African Journal of Education	1
Connection Science	1

truly rigorous methodology, such as the PRISMA framework. Given that educational research on ChatGPT is still in its infancy, the scarcity of experimental studies naturally limits the feasibility of meta-analyses. Similarly, bibliometric analysis requires a substantial body of work spanning several years and supported by a significant citation volume—conditions that this nascent field has yet to fulfill.

Bibliographic analysis reveals that the selected reviews were published across 23 scientific journals (see [Table 2](#)), with the highest number of contributions appearing in the following journals:

- Frontiers in Education (2);
- Contemporary Educational Technology (2);
- Computers and Education: Artificial Intelligence (2);
- International Journal of Learning, Teaching and Educational Research (2).

Most of the remaining journals published only a single article on the topic. The interdisciplinary nature of the subject, 'ChatGPT in education,' explains why systematic reviews are dispersed across a wide range of journals spanning diverse fields, including education, health, technology, and mathematics. Certain countries, notably Saudi Arabia, China, and Indonesia, are

Table 3. Geographical distribution of the included studies

Country	Number of publications
Saudi Arabia	5
China	4
Indonesia	3
United Arab Emirates	2
USA	2
Spain	2
Greece	2
Pakistan	1
Germany	1
Thailand	1
UK	1
Japan	1
Republic of Korea	1
South Africa	1

Table 4. Methodological characteristics of the included studies

Characteristic	Category	Frequency (n)
Databases searched	Scopus	18
	Google Scholar	16
	Web of Science	12
	ERIC	10
	ScienceDirect	8
	IEEE Xplore	7
	PubMed	6
	Others (JSTOR, ACM, SpringerLink, etc.)	5
Period covered	Recent period (2022 - 2024)	22
	Extended period (e.g., 2017 - 2023)	1
	Not specified / Up to search date	4
Quality assessment (QA)	Not mentioned / Not conducted	22
	Conducted (AMSTAR 2, Checklist, etc.)	5
Analytical approach	Qualitative	13
	Mixed (Qualitative & Quantitative)	13
	Quantitative only	1
Analysis method	Thematic analysis	20

overrepresented, indicating a regional concentration of systematic literature reviews on ChatGPT in education, particularly in Asia (see [Table 3](#)).

Methodological Characteristics of The Reviews (RQ2)

To address RQ2, we examined the methodological characteristics of the included systematic reviews, with a particular focus on the insights gained from analysis of searched databases and research scope, time period covered, inclusion and exclusion criteria and the AMSTAR 2 appraisal (see [Table 4](#)).

Searched databases and research scope

The examined reviews demonstrate a strong concentration on major academic databases, reflecting an effort to cover both the fields of education and technology. Scopus ($n = 18$) is the most frequently consulted database, followed by Google Scholar ($n = 16$), Web of Science ($n = 12$), and ERIC ($n = 10$). ScienceDirect ($n = 8$), IEEE Xplore ($n = 7$), and PubMed ($n = 6$) are also regularly used, while other databases such as JSTOR, ACM, and SpringerLink appear less frequently ($n = 5$). This combination suggests a commitment to embracing an interdisciplinary perspective, which is essential for a topic at the intersection of technology and pedagogy. The inclusion of Google Scholar, although sometimes criticized for its lack of rigor, is logical in an emerging field where grey literature and rapid publications are prevalent.

Time period covered

Regarding the period covered by these reviews, most focus on research published after 2022 (the year ChatGPT emerged). However, some reviews adopt a broader historical perspective by examining the evolution of AI use in education, covering studies published between 2017 and 2023 or between 2020 and 2024. While this allows for up-to-date data, it also indicates that research is still at an early stage. Indeed, studies often aim to quickly assess the impact of a new technology, which can influence both the nature and depth of the analyses produced.

Inclusion and exclusion criteria

All the analyzed reviews ($n = 27$) explicitly mention the inclusion and exclusion criteria, reflecting transparency in the systematic literature review process. The majority of the reviewed studies primarily rely on inclusion criteria, notably: articles written in English, availability of the full text, and a focus on the use of ChatGPT in the educational field. These reviews generally exclude gray literature (theses and dissertations, institutional reports, etc.), duplicates, as well as studies published before 2022 (the year ChatGPT emerged).

Methodological approaches and quality assessment

The research methodologies adopted in the reviews are mainly qualitative. More specifically, these reviews employ thematic analysis methods, narrative syntheses, or content analyses. Some reviews use mixed methods, combining qualitative and quantitative analyses by drawing on bibliometric data. Regarding the evaluation of methodological quality, while some reviews rely on checklists of methodological elements, more than half do not explicitly mention the indicators used to assess methodological quality. Thus, this shows that systematic reviews, even those using the PRISMA protocol, do not share common standards for evaluating methodological quality. The methodological assessment of the 27 reviews using the AMSTAR 2 tool revealed marked methodological heterogeneity in the existing literature. Indeed, only a minority of reviews ($n = 5$; 18.5%) achieved a moderate level of overall confidence. In contrast, the overwhelming majority ($n = 22$; 81.5%) were judged to have a low to critically low level of overall confidence. These findings highlight significant methodological shortcomings.

In particular, there is a lack of pre-registered protocols prior to initiating the data collection process (criterion 1), in databases such as PROSPERO or OSF. Moreover, most of the reviews did not provide detailed justifications for the studies excluded during the eligibility phase (criterion 3). Although the databases used (Scopus, Google Scholar, Web of Science, etc.) were explicitly mentioned, the authors did not always justify their selection, which limits transparency and increases methodological uncertainty (criterion 2). A further gap was identified in the assessment of the risk of bias (including selection, detection, and attrition bias, etc.) for the studies included in the analyzed systematic reviews (criterion 4).

Furthermore, the choice of an appropriate meta-analysis method was often absent or not justified in the statistical analyses of the results (criterion 5). Indeed, several systematic reviews conducted analyses without testing data heterogeneity. In addition, the consideration of the risk of bias when interpreting the results (criterion 6) was frequently overlooked. Finally, the absence of an assessment of publication bias, an additional AMSTAR 2 domain beyond the six criteria described above, suggests a possible overestimation of the benefits of AI, by masking the limitations or disadvantages of studies not published in scientific journals. Thus, according to AMSTAR 2, these methodological limitations mean that the overwhelming majority of systematic reviews on the use of ChatGPT in education achieve only a low or very low level of confidence.

Pedagogical Uses of ChatGPT in Education (RQ3)

In accordance with the analysis strategy described in Section *Analysis Method*, the synthesis draws on all included reviews while giving greater interpretative emphasis to findings consistently supported by reviews with higher methodological quality. To address RQ3, we analyzed the various pedagogical uses of ChatGPT identified in the literature, ranging from writing assistance to automated assessment and personalized learning. It should be noted that the systematic reviews included in this umbrella review

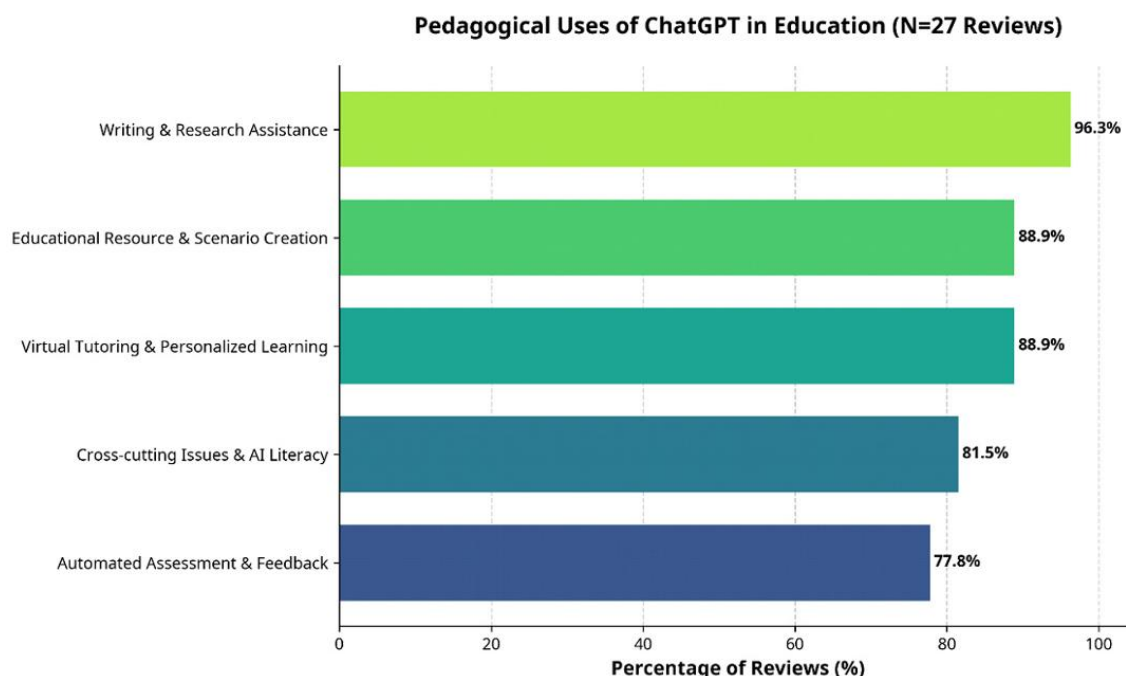


Figure 4. Pedagogical uses of ChatGPT in education (Source: Author's own elaboration)

generally did not specify the version(s) of ChatGPT examined. Consequently, potential differences in capabilities across ChatGPT versions could not be taken into account in the present synthesis. **Figure 4** shows that the majority of the analyzed reviews address research topics related to writing and research assistance (96.3%). It is also observed that the themes of educational resource and teaching scenario creation, as well as virtual tutoring and personalized learning, are significantly covered by the analyzed reviews (88.9%). Cross-cutting issues and AI literacy also appear highly represented (81.5%). However, automated assessment and feedback (77.8%) occupy a less prominent place, although they remain significant. This distribution suggests that current literature primarily prioritizes the two acts of the teaching process: course planning and learning assessment.

Writing and research assistance

The evidence synthesized from the reviews rated as having moderate methodological quality indicates that ChatGPT plays a central role in writing and research assistance (Park & Martin, 2024). These findings are further supported by the broader review literature. Several reviews report the central role of ChatGPT as a writing and research assistant. For instance, the study by Li et al. (2024) emphasizes that the tool can act as an information compiler and research support, while enabling idea generation, document structuring and organization. More specifically in the field of scientific and academic research, Al Faruq et al. (2023) consider that ChatGPT increases research efficiency by facilitating manuscript editing and academic publication. Explicitly, the study by Ofusori and Hendradi (2025) shows that ChatGPT is a sophisticated tool that allows researchers to tap into a vast amount of information across various fields. Beyond scientific research, ChatGPT plays a role in consolidating the writing skills of both teachers and learners by developing coherence, grammar, and writing style (Bašić et al., 2023). Furthermore, it facilitates the extraction of key points and the provision of feedback on references, thereby increasing researchers' efficiency (Bahrini et al., 2023; Morocco-Clarke et al., 2024).

Educational resource and teaching scenario creation

ChatGPT also supports the creation of educational content and case scenarios (Al Faruq et al., 2023; Fernández Cerero et al., 2025). For teachers, one concrete benefit stands out: less time spent preparing material tailored to diverse learning needs (Slamet & Basthomi, 2025). Teachers generate exercises, and assessment questions adapted to students' levels. Slamet and Basthomi (2025) also stipulate that ChatGPT can simulate real-life linguistic scenarios, allowing learners to practice and develop skills while facing concrete situations, improving not only linguistic proficiency but also cultural awareness. In the same vein, Heung and Chiu (2025) consider that ChatGPT enables group interactions in online learning environments while inspiring student creativity by suggesting topics for collaborative story writing and guiding scriptwriting.

Virtual tutoring and personalized learning

The studies conducted by Heung and Chiu (2025), and Park and Martin (2024) highlight that ChatGPT offers instant feedback and clarifications on a wide range of tasks, aiding in decision-making, problem-solving, conceptual knowledge building in programming, instructions and recommendations for science learning, and mathematics learning tasks (Jing et al., 2024; Wu et al., 2024). Fernández Cerero et al. (2025) emphasize that ChatGPT can act as a 24/7 virtual assistant, capable of answering students' questions outside of class hours, allowing them to write their research by interacting with the tool.

Cross-cutting issues and AI literacy

The cross-cutting issues related to Artificial Intelligence (AI) in education highlight the need to develop ethical and critical digital literacy among learners (Bettayeb et al., 2024). Mahsun et al. (2024) emphasize the importance of responsible AI tool usage, based on supervision, reflection on plagiarism, privacy protection, and the pursuit of equity. In the same vein, Dimeli and Kostas (2025) recommend raising students' awareness of the technical limitations of ChatGPT and promoting a critical and productive interaction with these tools. Li et al. (2023) highlight several transversal risks including plagiarism, hallucinations leading to fictitious references, algorithmic biases reproducing stereotypes related to gender, nationality, or culture, challenges related to data security, and risks related to the transparency and fairness of automated decisions in evaluation and administrative management processes.

Automated assessment and feedback

Slamet and Basthomi (2025) consider that ChatGPT supports the assessment of learning, automating the grading of learners' work and dynamically generating personalized assessment questions that adapt to students' individual needs. This is corroborated by Park and Martin (2024), whose review reports empirical evidence of ChatGPT grading student essays against a rubric and generating assessment tasks for classroom use. Li et al. (2024), and Al Faruq et al. (2023) show that ChatGPT can be used for automated writing assessment. While Al Faruq et al. (2023) highlight the pedagogical usefulness of ChatGPT and the reduction of teachers' workload, it is nonetheless recommended to monitor learners' work using plagiarism detection tools and to establish clear rules to ensure the reliability of assessments.

DISCUSSION

The results indicate that ChatGPT serves not only as a writing support tool but also as a research aid by facilitating bibliographic synthesis. This interpretation is consistent with the evidence synthesized in the review by Park and Martin (2024) and is further corroborated by more recent review studies (Huang et al., 2025; Milakis et al., 2025), which highlight ChatGPT's role as an editorial assistant that enhances writing productivity and efficiency. Although ChatGPT improves productivity and saves

time, some studies emphasize that the improved efficiency of the results obtained by this tool does not guarantee significant learning. Indeed, Morocco-Clarke et al. (2024) argue that reliance on ChatGPT tends to prevent students from focusing on critical analysis or can mask biased or fictitious citations, thus harming the rigor and reliability of research work.

The findings indicate that the integration of AI in education is not limited to the technical use of the tool, but is rooted in a structured framework of pedagogical design. This finding is also supported by the reviews with moderate overall confidence according to the AMSTAR 2 appraisal, particularly Heung and Chiu (2025), Fernández Cerero et al. (2025), and Park and Martin (2024), which emphasize the role of ChatGPT in instructional design and educational resource development. Accordingly, ChatGPT can be adopted as a pedagogical design tool for developing instructional scenarios adapted to learners' profiles, digital resources, case studies, and learning activities. For example, Madunić and Sovulj (2024) highlight the applications of ChatGPT in generating syllabi, lesson plans, learning objectives, and assessment grids. However, Yue et al. (2022) emphasize that although ChatGPT facilitates the automated generation of pedagogical scenarios, their educational value depends on alignment between learning activities (e.g., problem-solving and collaborative learning), pedagogical frameworks (e.g., constructivism, socio-constructivism, and problem-based learning), and technological tools. In other words, AI tools become pedagogically meaningful only when they are embedded within a coherent techno-pedagogical design that aligns learning objectives, instructional methods, and technological tools (Lebrun, 2011). Therefore, ChatGPT should be considered as a tool that supports pedagogical design rather than as a substitute for teachers' pedagogical expertise.

The findings further suggest that the most frequently described pedagogical uses in the analyzed reviews focus on virtual tutoring and automated assessment and feedback. These pedagogical applications are also consistently reported in the reviews with moderate overall confidence, particularly Heung and Chiu (2025), Fernández Cerero et al. (2025), and Park and Martin (2024).

These findings align with previous research (Farhood et al., 2025; Latif et al., 2026). Specifically, Latif et al. (2026) described artificial intelligence as a facilitator of personalized learning, adapting content to each learner's pace and needs. These systems offer constant support and immediate feedback, helping learners reduce their cognitive load.

Beyond technical uses, the integration of AI tools contributes to the transformation of the teacher's role. By automating some repetitive tasks, these tools enable teachers to focus more on developing learners' critical thinking and creativity, though their effective use requires mastery of the basis of prompt engineering. As noted by Madunić and Sovulj (2024), generating high-quality educational resources requires a strategy that defines an expert role and a well-defined context. Nevertheless, these outputs must always be reviewed, as they can sometimes contain inaccuracies that require validation by an expert.

The AMSTAR 2 appraisal reveals a methodological fragility within the field. This immaturity, characterized by a notable absence of pre-registered protocols and a persistent conflation between types of literature reviews (Arksey & O'Malley, 2005), weakens the reliability of current data. Transparency suffers too: the lack of explicit justification for exclusion criteria undermines reproducibility. And the failure to consider publication bias points to a broader pattern — a positive bias toward technological innovations, which may lead to an underestimation of risks—specifically hallucinations and cognitive dependency on AI (Li et al., 2023). This underscores the need for regular verification of the conclusions drawn from the current literature.

CONCLUSION AND IMPLICATIONS

This umbrella review aimed to explore research trends related to the use of ChatGPT in education. More specifically, it sought to address three research questions concerning the bibliographic characteristics of the analyzed literature reviews, their methodological approaches, and the various educational uses of ChatGPT in education. The results indicate a rapid growth in publications since the emergence of ChatGPT in 2022. In particular, systematic reviews are predominant, whereas meta-analyses and bibliometric analyses remain underexplored. From a methodological perspective, the analyzed reviews primarily rely on recognized databases such as Scopus, Web of Science, ScienceDirect and ERIC. Moreover, while most reviews explicitly describe inclusion and exclusion criteria, methodological quality assessment is not explicitly mentioned. Regarding educational applications, ChatGPT is used as a tool for writing and research, instructional design, virtual tutoring or personalized learning and automated assessment. Based on these findings, several operational recommendations are formulated for researchers and practitioners. Beyond synthesizing the existing review literature, this umbrella review also provides implications for educational research and practice by identifying methodological priorities and evidence-informed directions for the pedagogically meaningful integration of ChatGPT into education.

For researchers, future systematic reviews on ChatGPT should take into account the evaluation of the methodological quality of the studies analyzed, in order to ensure the reliability of the results obtained. Notably, we recommend to researchers that register their protocol in advance in recognized databases such as PROSPERO or OSF, so as to ensure transparency and reproducibility. Future studies should also broaden their geographical and linguistic scope in order to produce findings with stronger generalizability. Additionally, future systematic reviews could mobilize techno-pedagogical frameworks such as the TPACK model framework to better evaluate the extent to which ChatGPT transforms pedagogical engineering rather than merely substituting traditional practices. For teachers and educational practitioners, the use of ChatGPT should not be considered as a simple substitution of traditional tools, but rather as an opportunity to intentionally redesign instructional scenarios ensuring techno-pedagogical alignment between learning objectives, pedagogical strategies, and the affordances of the AI tool. Concretely, teachers are encouraged to design prompting resources that develop learners' critical thinking rather than reinforcing passive reliance on AI-generated responses. Furthermore, the development of AI literacy including awareness of algorithmic limitations, hallucinations, and ethical issues such as plagiarism and data privacy should be explicitly embedded into courses instead of being treated as a supplementary aspect.

This study has limitations, which may serve as a basis for future research. First, the search process was limited to databases such as Scopus, Web of Science, ScienceDirect, and ERIC. Second, during the article selection phase (following the PRISMA approach), only studies written in English or French were included. In addition to that, a formal overlap analysis (e.g., Corrected Covered Area) could not be conducted because a substantial number of the included reviews do not report the complete list of their primary studies, either omitting this information entirely or providing it only in a partial or aggregated form. This lack of transparency in several of the primary reviews made a reliable overlap index methodologically unfeasible to compute. Consequently, although care was taken during the study selection process to minimize substantial content overlap, some duplication of primary studies across the included reviews cannot be entirely excluded.

Furthermore, this study was restricted to journal articles. Future research could broaden this scope by including conference proceedings and books so as to provide a deeper analysis of research on the use of ChatGPT in education.

Funding: No funding source is reported for this study.

Ethical statement: The author stated that the study is a literature-based umbrella review that did not involve human participants, patient data, or animal subjects. Consequently, approval from an institutional ethics committee was not required. No sensitive or confidential personal data were collected or processed in this research.

AI statement: The author stated that, during the preparation of this manuscript, the author used Claude Sonnet 5 (Anthropic) to check grammar and improve certain phrasings for greater fluency and naturalness in English. The tool was not used to generate scientific content, ideas, analyses, or conclusions. Even for these language edits, the author carefully reviewed and validated each suggested change before incorporating it into the manuscript. The author takes full responsibility for the accuracy, integrity, and originality of the final content.

Declaration of interest: No conflict of interest is declared by the author.

Data sharing statement: Data supporting the findings and conclusions are available upon request from the author.

REFERENCES

- Abdellatif, M., Dadam, M. N., Vu, N. T., Nam, N. H., Hoan, N. Q., Taoube, Z., ... & Huy, N. T. (2025). A step-by-step guide for conducting an umbrella review. *Tropical Medicine and Health*, 53(1), Article 134. <https://doi.org/10.1186/s41182-025-00764-y>
- Arksey, H., & O'Malley, L. (2005). Scoping studies: Towards a methodological framework. *International Journal of Social Research Methodology*, 8(1), 19-32. <https://doi.org/10.1080/1364557032000119616>
- Bahrini, A., Khamoshifar, M., Abbasimehr, H., Riggs, R. J., Esmaeili, M., Majdabadkohne, R. M., & Pasehvar, M. (2023, April). ChatGPT: Applications, opportunities, and threats. In *2023 systems and information engineering design symposium (SIEDS)* (pp. 274-279). IEEE. <https://doi.org/10.1109/SIEDS58326.2023.10137850>
- Baig, M. I., & Yadegaridehkordi, E. (2024). ChatGPT in the higher education: A systematic literature review and research challenges. *International Journal of Educational Research*, 127, Article 102411. <https://doi.org/10.1016/j.ijer.2024.102411>
- Bašić, Ž., Banovac, A., Kružić, I., & Jerković, I. (2023). ChatGPT-3.5 as writing assistance in students' essays. *Humanities and Social Sciences Communications*, 10(1), Article 750. <https://doi.org/10.1057/s41599-023-02269-7>
- Bernard, F., & Fluckiger, C. (2019). Innovation technologique, innovation pédagogique. *Spirale-Revue de Recherches en Education*, 63(1), 3-10. <https://doi.org/10.3917/spir.063.0003>
- Conn, V. S., & Coon Sells, T. G. (2014). WJNR welcomes umbrella reviews. *Western Journal of Nursing Research*, 36(2), 147-151. <https://doi.org/10.1177/0193945913506968>
- de la Higuera, C., & Iyer, J. (2024). *AI for teachers: An open textbook*. Pressbooks. <https://pressbooks.pub/aiforteachers/>
- Edhuman. (2024). *L'IA au service de l'éducation : Guide pratique pour enseignants et équipes pédagogiques [AI in Education: A practical guide for teachers and teaching teams]*. <https://www.edhuman.fr/ia-education-guide>
- European Parliament. (2020, August 27). *What is artificial intelligence and how is it used?* <https://www.europarl.europa.eu/topics/en/article/20200827STO85804>
- Farhood, H., Nyden, M., Beheshti, A., & Muller, S. (2025). Artificial intelligence-based personalised learning in education: a systematic literature review. *Discover Artificial Intelligence*, 5(1), Article 331. <https://doi.org/10.1007/s44163-025-00598-x>
- Fauvet, C., & Gatignol, P. (2022). La revue systématique de la littérature et la méta-analyse [Systematic literature review and meta-analysis]. In S. Borel, P. Gatignol, A. Gros, & T. M. Tran (Éds.), *Manuel de recherche en orthophonie : Fondements de la recherche en orthophonie, formation initiale et continue, toutes les UE Recherche* (pp.149-161). De Boeck Supérieur. <https://shs.cairn.info/manuel-de-recherche-en-orthophonie--9782807335226-page-149?lang=fr>
- Hasni, A., Bousadra, F., Belletête, V., Benabdallah, A., Nicole, M. C., & Dumais, N. (2016). Trends in research on project-based science and technology teaching and learning at K-12 levels: A systematic review. *Studies in Science Education*, 52(2), 199-231. <https://doi.org/10.1080/03057267.2016.1226573>
- Huang, R., Yin, Y., Zhou, N., & Lang, F. (2025). Artificial intelligence in K-12 education: An umbrella review. *Computers and Education: Artificial Intelligence*, 10, Article 100519. <https://doi.org/10.1016/j.caeai.2025.100519>
- Javaid, M., Haleem, A., Singh, R. P., Khan, S., & Khan, I. H. (2023). *Unlocking the opportunities through ChatGPT tool towards ameliorating the education system. BenchCouncil Transactions on Benchmarks, Standards and Evaluations*, 3(2), Article 100115. <https://doi.org/10.1016/j.tbench.2023.100115>

- Jing, Y., Wang, H., Chen, X., & Wang, C. (2024). What factors will affect the effectiveness of using ChatGPT to solve programming problems? A quasi-experimental study. *Humanities and Social Sciences Communications*, 11(1), 1-12. <https://doi.org/10.1057/s41599-024-02751-w>
- Jmoula, L. (2023). Enseigner et apprendre à l'ère de ChatGPT : des pistes issues d'une recension des écrits [Teaching and learning in the age of ChatGPT: insights from a review of the literature]. *Journal of Information Sciences*, 22(2), 26-40. <https://doi.org/10.34874/IMIST.PRSM/jis-v22i2.41122>
- Joyce, K., Smith-Doerr, L., Alegria, S., Bell, S., Cruz, T., Hoffman, S. G., ... & Shestakofsky, B. (2021). Toward a sociology of artificial intelligence: A call for research on inequalities and structural change. *Socius: Sociological Research for a Dynamic World*, 7. <https://doi.org/10.1177/2378023121999581>
- Latif, E., Liu, V., & Zhai, X. (2026). A systematic review of intelligent and robot tutoring systems: evolution, pedagogical design, and AI-driven classification. *Smart Learning Environments*, 13(1), Article 1. <https://doi.org/10.1186/s40561-025-00427-9>
- Lebrun, M. (2011). Impacts des TIC sur la qualité des apprentissages des étudiants et le développement professionnel des enseignants: Vers une approche systémique [Impacts of ICT on the quality of student learning and the professional development of teachers: Towards a systemic approach]. *Sciences et Technologies de l'Information et de la Communication pour l'Éducation et la Formation*, 18, 287-316. <https://doi.org/10.3406/stice.2011.1028>
- Madunić, J., & Sovulj, M. (2024). Application of ChatGPT in information literacy instructional design. *Publications*, 12(2), Article 11. <https://doi.org/10.3390/publications12020011>
- McCarthy, J. (2007). *What is artificial intelligence?* Stanford University. <http://www-formal.stanford.edu/jmc/whatisai.pdf>
- Milakis, E. D., Argyrakou, C. C., Melidis, A., & Vrettaros, J. (2025). ChatGPT and AI chatbots in education: An umbrella review of systematic reviews, scoping reviews, and meta-analyses. *International Journal of Education and Information Technologies*, 19, 100-119. <https://doi.org/10.46300/9109.2025.19.11>
- Milmo, D. (2023, February 2). ChatGPT reaches 100 million users two months after launch. *The Guardian*. <https://www.theguardian.com/technology/2023/feb/02/chatgpt-100-million-users-open-ai-fastest-growing-app>
- Morocco-Clarke, A., Sodangi, F. A., & Momodu, F. (2024). The implications and effects of ChatGPT on academic scholarship and authorship: a death knell for original academic publications? *Information & Communications Technology Law*, 33(1), 21-41. <https://doi.org/10.1080/13600834.2023.2239623>
- Pan, C., Banerjee, J. S., De, D., Sarigiannidis, P., Chakraborty, A., & Bhattacharyya, S. (2023, February). ChatGPT: A OpenAI platform for society 5.0. In *Doctoral symposium on human centered computing* (pp. 384-397). Springer Nature Singapore. https://doi.org/10.1007/978-981-99-3478-2_33
- Piekoszewski-Cuq, P. (2024). *Intelligence artificielle : Guide pratique pour les enseignants* [Artificial intelligence: A practical guide for teachers]. https://pedagogie.ac-guadeloupe.fr/sites/default/files/2025-02/guide_ia_enseignants_1.pdf
- Wu, T. T., Lee, H. Y., Li, P. H., Huang, C. N., & Huang, Y. M. (2024). Promoting self-regulation progress and knowledge construction in blended learning via ChatGPT-based learning aid. *Journal of Educational Computing Research*, 61(8), 1539-1567. <https://doi.org/10.1177/07356331241236363>
- Yue, M., Jong, M. S. Y., & Dai, Y. (2022). Pedagogical design of K-12 artificial intelligence education: A systematic review. *Sustainability*, 14(23), Article 15620. <https://doi.org/10.3390/su142315620>

APPENDIX

Table A1. List of included studies

Adarkwah, M. A., Badu, S. A., Osei, E. A., Adu-Gyamfi, E., Odame, J., & Schneider, K. (2025). ChatGPT in healthcare education: A double-edged sword of trends, challenges, and opportunities. <i>Discover Education</i> , 4, Article 14. https://doi.org/10.1007/s44217-024-00393-3
Alibrahim, A. A. (2024). Is ChatGPT reliable in education? <i>South African Journal of Education</i> , 44(4), 1-9. https://doi.org/10.15700/saje.v44n4a2557
Al Faruq, M. S. S., Sunoko, A., Ibda, H., & Wahyudi, K. (2023). Digital learning management using OpenAI ChatGPT: A systematic literature review. <i>International Journal of Learning, Teaching and Educational Research</i> , 22(12), 21-41. https://doi.org/10.26803/ijlter.22.12.2
Ansari, A. N., Ahmad, S., & Bhutta, S. M. (2024). Mapping the global evidence around the use of ChatGPT in higher education: A systematic scoping review. <i>Education and Information Technologies</i> , 29, 11281-11321. https://doi.org/10.1007/s10639-023-12223-4
Atiyah, A. N. (2023). Tutoring and education: Identifying the opportunities and challenges of ChatGPT among students. <i>Communications in Mathematics and Applications</i> , 14(5), 1799-1813. https://doi.org/10.26713/cma.v14i5.2687
Bettayeb, A. M., Abu Talib, M., Sobhe Altayasinah, A. Z., & Dakalbab, F. (2024, July). Exploring the impact of ChatGPT: Conversational AI in education. <i>Frontiers in Education</i> , 9, Article 137979. https://doi.org/10.3389/feduc.2024.1379796
Chytas, D., Nossios, G., Paraskevas, G., Vasiliadis, A. V., Giovanidis, G., & Troupis, T. (2025). Can ChatGPT play a significant role in anatomy education? A scoping review. <i>Morphologie</i> , 109(365), Article 100949. https://doi.org/10.1016/j.morpho.2025.100949
Crompton, H., & Burke, D. (2024). The educational affordances and challenges of ChatGPT: State of the field. <i>TechTrends</i> , 68, 380-392. https://doi.org/10.1007/s11528-024-00939-0
Deng, R., Jiang, M., Yu, X., Lu, Y., & Liu, S. (2024). Does ChatGPT enhance student learning? A systematic review and meta-analysis of experimental studies. <i>Computers & Education</i> , 227, Article 105224. https://doi.org/10.1016/j.compedu.2024.105224
Dimeli, M., & Kostas, A. (2025). The role of ChatGPT in education: Applications, challenges: Insights from a systematic review. <i>Journal of Information Technology Education: Research</i> , 24(1), Article 2. https://doi.org/10.28945/5422
Evangelista, E. D. L. (2025). Ensuring academic integrity in the age of ChatGPT: Rethinking exam design, assessment strategies, and ethical AI policies in higher education. <i>Contemporary Educational Technology</i> , 17(1), Article ep559. https://doi.org/10.30935/cedtech/15775
Faisal, E. (2024). Unlock the potential for Saudi Arabian higher education: A systematic review of the benefits of ChatGPT. <i>Frontiers in Education</i> , 9, Article 1325601. https://doi.org/10.3389/feduc.2024.1325601
Fernández Cerero, J., Montenegro Rueda, M., Román Graván, P., & Fernández Batanero, J. M. (2025). ChatGPT as a digital tool in the transformation of digital teaching competence: A systematic review. <i>Technologies</i> , 13(5), Article 205. https://doi.org/10.3390/technologies13050205
Gunawan, J., Aunguroch, Y., & Montayre, J. (2024). ChatGPT integration within nursing education and its implications for nursing students: A systematic review and text network analysis. <i>Nurse Education Today</i> , 141, Article 106323. https://doi.org/10.1016/j.nedt.2024.106323
Heung, Y. M. E., & Chiu, T. K. (2025). How ChatGPT impacts student engagement from a systematic review and meta-analysis study. <i>Computers and Education: Artificial Intelligence</i> , 8, Article 100361. https://doi.org/10.1016/j.caeai.2025.100361
Imran, M., & Almusharraf, N. (2023). Analyzing the role of ChatGPT as a writing assistant at higher education level: A systematic review of the literature. <i>Contemporary Educational Technology</i> , 15(4), Article ep464. https://doi.org/10.30935/cedtech/13605
Khan, N., Khan, Z., Koubaa, A., Khan, M. K., & bin Salleh, R. (2024). Global insights and the impact of generative AI-ChatGPT on multidisciplinary: A systematic review and bibliometric analysis. <i>Connection Science</i> , 36(1), Article 2353630. https://doi.org/10.1080/09540091.2024.2353630
Li, M., Enkhtur, A., Cheng, F., & Yamamoto, B. A. (2023). Ethical implications of ChatGPT in higher education: A scoping review. <i>arXiv preprint arXiv:2311.14378</i> . https://doi.org/10.48550/arXiv.2311.14378
Li, B., Lowell, V. L., Wang, C., & Li, X. (2024). 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. <i>Computers and Education: Artificial Intelligence</i> , 7, Article 100266. https://doi.org/10.1016/j.caeai.2024.100266
Mahsun, M., Ali, M., Ekaningrum, I. R., & Ibda, H. (2024). Trend of using ChatGPT in learning process and character education: A systematic literature review. <i>International Journal of Learning, Teaching and Educational Research</i> , 23(5), 387-402. https://doi.org/10.26803/ijlter.23.5.20
Newton, P., & Xiromeriti, M. (2024). ChatGPT performance on multiple choice question examinations in higher education: A pragmatic scoping review. <i>Assessment & Evaluation in Higher Education</i> , 49(6), 781-798. https://doi.org/10.1080/02602938.2023.2299059
Ofusori, L. O., & Hendradi, R. (2025). ChatGPT and its impact on students assessment practices in the higher educational sector: A systematic review. <i>Journal of Information Systems Engineering and Business Intelligence</i> , 11(1), 65-78. https://doi.org/10.20473/jisebi.11.1.65-78
Park, H. K., & Martin, S. N. (2024). Exploring the role of ChatGPT in science education for Asia-Pacific and beyond: A systematic review. <i>Asia-Pacific Science Education</i> , 10(2), 233-264. https://doi.org/10.1163/23641177-bja10079
Slamet, J., & Basthomi, Y. (2025). Examining the challenges and opportunities of ChatGPT in EFL education: A systematic literature review. <i>Journal of University Teaching and Learning Practice</i> , 22(2), 1-26. https://doi.org/10.53761/deezkh88
Teng, M. F. (2024). A systematic review of ChatGPT for English as a foreign language writing: Opportunities, challenges, and recommendations. <i>International Journal of TESOL Studies</i> , 6(3), 36-57. https://doi.org/10.58304/ijts.20240304
Xu, X., Chen, Y., & Miao, J. (2024). Opportunities, challenges, and future directions of large language models, including ChatGPT in medical education: A systematic scoping review. <i>Journal of Educational Evaluation for Health Professions</i> , 21, Article 6. https://doi.org/10.3352/jeehp.2024.21.6
Zhang, P., & Tur, G. (2024). A systematic review of ChatGPT use in K-12 education. <i>European Journal of Education</i> , 59(4), Article e12599. https://doi.org/10.1111/ejed.12599