Zhang et al. *Research and Practice in Technology Enhanced Learning* (2026) 21:34
https://doi.org/10.58459/rptel.2026.21034

Research and Practice in Technology Enhanced Learning

RESEARCH Free and Open Access

Generative artificial intelligence in K-12 education: A systematic review

Tianyu Zhang *, Yiu Chi Lai and Philip Leung Ho Yu

*Correspondence:
s1146795@s.eduhk.hk
Department of Mathematics and Information Technology,
The Education University of Hong Kong,
10 Lo Ping Road, New Territories, Hong Kong S.A.R.
Full list of author information is available at the end of the article

Abstract

With the continuous innovation of deep learning algorithms, Generative Artificial Intelligence (GenAI) technology is rapidly developing globally and gradually expanding its application scenarios in multiple fields, especially in education. Considering the novelty of this field, there is currently a scarcity of comprehensive research on GenAI in K-12 education. Therefore, this systematic review aims to reveal the application trends, teaching themes, tool adoption, research methods, challenges, and advantages of generative artificial intelligence in K-12 education through the in-depth analysis of 45 studies between 2020 and 2024, providing theoretical and empirical support for future research and practice in this field. Our thematic analysis results indicate that GenAI tools can significantly improve students' academic performance and cognitive abilities, enhance their learning motivation, and thus promote the development of personalized learning. However, using these tools also brings a series of challenges, including misleading or erroneous content generation, difficulty in understanding technology, students' dependence on technology, and privacy infringement. In addition, the shortcomings demonstrated by educators in terms of AI literacy emphasize the necessity for relevant educational institutions to organize targeted AI literacy training. Finally, given that we are currently in the early stages of developing generative artificial intelligence, most existing empirical research has focused on the short-term impact of GenAI tools on K-12 education. Future research will incorporate more longitudinal studies to systematically evaluate the long-term and deep implications of GenAI in education.

Keywords: Generative artificial intelligence, K-12 education, Large language models, AI literacy, ChatGPT, Learning outcomes, Systematic review

Introduction

Ever since Alan Turing initially expounded the possibility of "machine thinking" and the concept of the "Turing Test" in 1950, the research in artificial intelligence (AI) has progressed across diverse domains, engendering a burgeoning corpus of scholarly

discourse (Zhang & Aslan, 2021). Artificial intelligence aims to build machines and algorithms that can simulate human intelligence and technology to perform various sophisticated tasks related to the development of human society (Liu et al., 2018). These tasks include but are not limited to decision-making, image recognition, linguistic analysis, and content generation. Integrating artificial intelligence into the education system represents a paradigm shift in teaching and learning practices (Chen et al., 2022).

Artificial intelligence in education (AIED) refers to using artificial intelligence technology in both traditional classroom settings and remote teaching environments, aiming to imitate teachers' perceptions and instructional strategizing methodologies (Roll & Wylie, 2016). Since the advent of artificial intelligence, researchers have been steadfastly dedicated to developing intelligent tutoring systems endowed with adaptive capabilities for guiding human learners (VanLehn, 2011). The COVID-19 pandemic has significantly accelerated the integration of AI in education, profoundly impacting the conventional educational system (Kang, 2021; Pantelimon et al., 2021; Xie et al., 2020). Generative artificial intelligence (GenAI) refers to a subcategory of artificial intelligence technology and methodology that focuses on the synthesis of novel and creative content (Banh & Strobel, 2023), encompassing a broad spectrum of modalities such as text content, audio, visual imagery, and video materials.

There is a rapid proliferation of diverse generative artificial intelligence tools in contemporary digital education. ChatGPT and GPT-4, renowned for their proficiency in natural language processing (Qin et al., 2023), can provide students with instantaneous access to an array of educational materials. The available resources encompass creative course content (Rüdian & Pinkwart, 2023), tailored study plans (Kostka & Toncelli, 2023; van den Berg & du Plessis, 2023), comprehensive exercise solutions (Farooq & Anwar, 2023; Fergus et al., 2023), concise summarizing knowledge graphs (Vassiliou et al., 2023), effective exam preparation strategies, and exam questions (Aboalela, 2023). The visual images created by Midjourney can stimulate students' imagination and creativity (Chiu, 2024), and Midjourney can design illustrations of fairy tales for young children (Ruskov, 2023), combined with the use of ChatGPT to create interactive children's books (Kim & Kim, 2023). The D-ID website, which utilizes virtual avatars using existing photos, enhances student engagement through teachers' virtual avatars in learning (Alam & Mohanty, 2022). GitHub Copilot is an AI-powered code completion tool, co-engineered by GitHub and OpenAI, which has demonstrated its efficacy as an adjunct educational aid within programming curricula (Puryear & Sprint, 2022; Wermelinger, 2023). GenAI tools centered on ChatGPT facilitate personalized and interactive learning for students and alleviate the burden of course material preparation for teachers.

Moreover, prior research endeavors have attempted to undertake comprehensive literature reviews from diverse perspectives to elucidate the concept of artificial

intelligence applications in education. For example, the literature review authored by Baytak (2023) investigated the reception and assimilation of large language models (LLMs), specifically ChatGPT and Google Bard, within pedagogical contexts. Imran and Almusharraf (2023) executed a systematic literature review of 30 studies on ChatGPT in higher education writing. Their research aimed to analyze the implications of employing ChatGPT within university-level writing programs and to examine the evolving landscape of instruction in writing competencies in academic settings. Baidoo-Anu and Ansah's (2023) literature review synthesized research to unveil the potential advantages and identified the limitations associated with deploying ChatGPT and correlating generative AI technologies in educational settings. Rizvi et al. (2023) reviewed 28 literature from 2019 to 2022 on the application of artificial intelligence teaching in primary and secondary schools, and based on limited empirical research, it was found that enhancing the instruction and mastering of artificial intelligence in primary and secondary education settings may be achieved by shifting a pedagogical framework that prioritizes the student experience, incorporating teaching strategies that are responsive to learners' contextual needs, and establishing comprehensive metrics for evaluating the effectiveness of AI educational initiatives. The literature on artificial intelligence and education was reviewed and analyzed by Zhou et al. (2020), leading to the development of an explicit conceptual framework for implementing artificial intelligence in K-12 AI education.

Although these reviews comprehensively summarized GenAI's numerous achievements in the field of education, some of them primarily focused on higher education rather than K-12 education (Baytak, 2023; Imran & Almusharraf, 2023), or the selection of GenAI tools is limited (Baidoo-Anu & Ansah, 2023). Additionally, Zhou et al. (2020) provided limited support for K-12 AI literacy design tools and courses or predominantly concentrated on bibliometric analysis within this domain (Bahroun et al., 2023; Bozkurt, 2023). Given the current lack of literature reviews specifically examining the implementation of various generative artificial intelligence systems in K-12 education, a more comprehensive systematic literature review (SLR) is necessary to address the following research questions effectively:

- RQ1 - What is the current overview of the research on GenAI in K-12 education?
- RQ2 - What are the significant advantages of utilizing GenAI tools in K-12 education?
- RQ3 - What challenges are associated with integrating GenAI in K-12 education?

This study aimed to conduct a systematic review by collating research on the application of generative artificial intelligence in K-12 education from 2020 to 2024 to address the aforementioned research inquiries. Systematic reviews enable individuals to independently

assess the relevance and robustness of the presented evidence and evaluate the integrity of the employed review process. Their comprehensive search strategy ensures the inclusion of a diverse range of studies, surpassing those highly cited or widely recognized (Bearman et al., 2012). By employing this mode of expression, readers are equipped with the means to acquire a comprehensive understanding of the existing knowledge about a subject, assess the credibility of study outcomes, and discern the collective impact of research endeavors.

Methodology

This systematic literature review adopted the preferred reporting items for systematic reviews and meta-analyses (PRISMA) 2020 structured framework (Page et al., 2021), intended to comprehensively and systematically summarize and evaluate the evidence of all relevant studies in a specific research field while addressing research questions. This study used reference management, data analysis, and visualization software, including RefWorks, MAXQDA 24, and VOSviewer. The following subsections provide a detailed description of the methodology used in this study.

Search databases

This study retrieved and selected articles from the following digital databases:

- Education Resources Information Centre (ERIC)
- Web of Science (WoS)
- Scopus
- IEEE Xplore Digital Library
- Taylor & Francis
- Wiley

Inclusion and exclusion criteria

The establishment of stringent inclusion and exclusion criteria is imperative to ensure the attainment of compelling analytical results.

The specified inclusion criteria are as follows:

- Articles' content is in English
- Published from January 2020 to April 2024
- Complete journal and conference articles
- Research on the application of generative AI systems or software in K-12 education
- Articles that present research findings based on empirical investigation

The specified exclusion criteria are as follows:

- Remove duplicate articles
- Remove conference articles with only abstracts

- Remove articles that belong to the field of AI but not generative AI
- Remove review-related articles
- Remove articles not available through open-access

Search strategy

Inclusion criteria were initially applied to the title, keywords, and abstract for relevance, and articles within the topic scope were subsequently subjected to a full-text review for further consideration. This study searched the selected digital databases (ERIC, WoS, Scopus, IEEE Xplore, Taylor & Francis, and Wiley) for all relevant articles that met the inclusion criteria from January 2020 to April 2024. The OR operator is utilized to retrieve articles about any given topic located on either side of the OR. Conversely, the AND operator is employed to extract articles that simultaneously mention two search topics. The combinations of search keywords used by the four databases (ERIC, WoS, Scopus, and IEEE Xplore) are as follows:

(generative **AND** artificial **AND** intelligence **OR** generative **AND** ai **OR** gai **OR** genai **OR** chatgpt) **AND** (k12 **OR** k-12 **OR** elementary **AND** education **OR** secondary **AND** education **OR** primary **AND** school **OR** elementary **AND** school **OR** middle **AND** school **OR** secondary **AND** school **OR** high **AND** school)

If the above search keywords were used in the two databases (Taylor & Francis and Wiley), tens of thousands of unrelated search results would be generated. Therefore, a combination of modified search keywords as shown below was used to narrow the screening scope and shorten review time:

(generative artificial intelligence **OR** generative ai **OR** gai **OR** genai **OR** chatgpt) **AND** (k12 **OR** k-12 **OR** elementary education **OR** secondary education **OR** primary school **OR** elementary school **OR** middle school **OR** secondary school **OR** high school)

Data extraction

The distribution of search results in the digital databases used in the data extraction process is as follows: ERIC (n = 39), WoS (n = 2539), Scopus (n = 457), IEEE Xplore (n = 1356), Taylor & Francis (n = 389), and Wiley (n = 476). All the displayed results were imported into RefWorks, and 426 duplicates were removed using the remove duplicates feature of the literature management software. In the screening stage for preliminary viewing, the remaining 4830 articles were reviewed for their titles and abstracts to determine whether they were excluded. However, the vast majority of these articles did not meet the inclusion criteria for all items, especially some that only focused on generative artificial intelligence technology rather than combining it with education; a portion of the articles consisted of review articles, while others were situated within the domain of higher education research; there are articles that, despite being located within the K-12 education sphere, involve the

active participation of educators rather than K-12 students. Moreover, during the screening phase, the authors found only a few articles that addressed preschool education, and none included preschool learners (Luo et al., 2023; Rakap, 2023; Su & Yang, 2023). Therefore, this systematic review did not include articles on preschool education. 4705 articles were excluded in the screening stage, and 125 articles were retained.

In the eligibility stage, the authors scanned the full text of 125 articles and excluded 80 of them based on four reasons for exclusion. The following are examples of exclusions in the eligibility stage. The online AI graphical programming platform designed by Aung et al. (2022) did not involve generative AI technology. Javier and Moorhouse (2024) conducted a study on English learning of K-12 participants using ChatGPT but did not present sufficient data analysis process and research findings. Leng et al. (2024) established a dataset to assist children's language acquisition that did not conduct empirical research involving K-12 students. In the included stage, 45 articles met the inclusion criteria (see Figure 1 below).

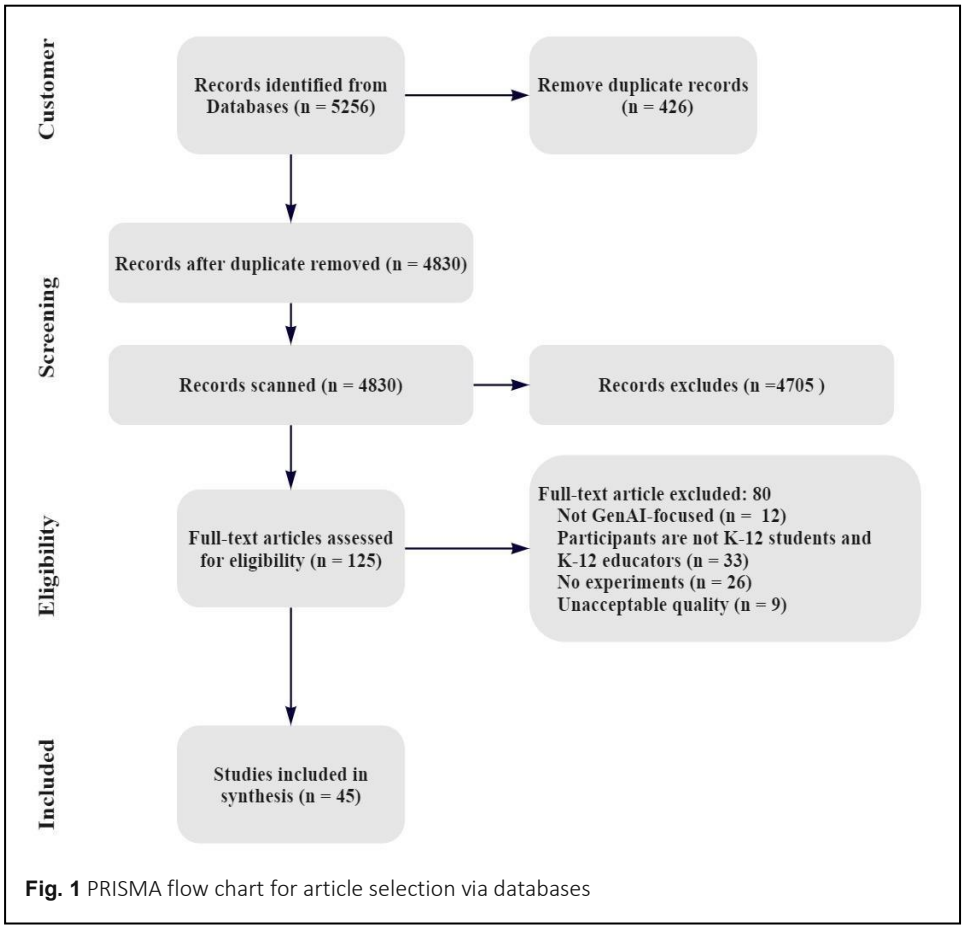

Fig. 1 PRISMA flow chart for article selection via databases

Results

In this section, we comprehensively reviewed the entire text of the selected articles, extracted data from the selected articles based on the research questions we want to address, and analyzed the results to provide as objective and fair an explanation and summary as possible. As the application of generative artificial intelligence in K-12 education is an emerging field, we did not impose restrictions on conference articles, but rigorously reviewed the richness of the content and persuasiveness of the evidence provided.

RQ1 - What is the current overview of the research on applying GenAI in K-12 education?

To comprehensively address the current research landscape of GenAI applications in K-12 education, this article concisely presented the information about Publication year, Publication type, Context, Subject, Research sample size, Age and grade level, and Research methodology.

Research overview of selected articles

According to Table 1, while this investigation's primary objective revolved around identifying scholarly literature from January 2020 to April 2024, no records emerged during 2020. Notably, more than half of the publications were published in 2023 (n = 31, 68.9%), followed by 2024 (n = 10, 22.2%), 2021 (n = 3, .67%), and 2022 (n = 1, 2.2%). The predominant publication format was journal articles (n = 30, 66.7%), followed by conference papers (n = 15, 33.3%). Research was most commonly focused on Asia (n = 18), followed by North America (n = 11), Europe (n = 10), South America (n = 3), Middle East (n = 2), and Africa (n = 1). Regarding the research methodology of the selected articles, the most commonly employed research methodology was mixed methods (n = 22), followed by qualitative methods (n = 13) and quantitative methods (n = 10).

Table 1 and 2 offer a comprehensive and systematic overview of all literature, including contexts, topics, sample sizes, participants' ages, research methodologies, instruments, and concise summaries. Most studies during this period focused on secondary education (n = 25), followed by primary education (n = 13), primary and secondary education (n = 5), and K-12 education (n = 2). All studies provided contextual information regarding the country or region in which they were performed. In addition to K-12 students (n = 29) and educators (n = 11), a few studies (n = 5) used GenAI tools as research subjects. Some studies (n = 14) did not specify the age or grade of the participants. These studies were conducted in different cultural, linguistic, or geographical contexts. In the disciplines or research areas of the selected studies, STEM (Science, n = 9; Technology, n = 7; Mathematics, n = 6) was the field that researchers focused on the most, followed by English

(n = 8), art (n = 4), history (n = 3), school education (n = 3), STEAM with art added to STEM (n = 2), religion (n = 1), literacy (n = 1), and one study involved multiple disciplines (Bekeš & Galzina, 2023). The table information provides the application effectiveness of GenAI tools and corresponding teaching methods for this review, thereby elucidating substantive contributions, discernible patterns, and trends in the research field.

Table 1 Frequencies and percentages of literature characteristics

Category	Characteristic	Frequency	Percentage (%)
Publication year	Total	45	100
	2021	3	6.7
	2022	1	2.2
	2023	31	68.9
	2024	10	22.2
Publication type	Total	45	100
	Journal article	30	66.7
	Conference paper	15	33.3
Research methodology	Total	45	100
	Quantitative	13	28.9
	Qualitative	11	24.4
	Mixed	21	46.7
Region of studies	Total	45	100
	Asia	*18*	*40*
	Hong Kong S.A.R.	6	13.3
	South Korea	4	8.9
	Taiwan, China	3	6.7
	China	2	4.5
	Singapore	1	2.2
	Indonesia	1	2.2
	Thailand	1	2.2
	North America	*11*	*24.4*
	USA	10	22.2
	Canada	1	2.2
	Europe	*10*	*22.2*
	German	3	6.7
	Finland	2	4.5
	Italy	2	4.4
	Ireland	1	2.2
	Spain	1	2.2
	Croatia	1	2.2
	South America	*3*	*6.7*
	Uruguay	1	2.3
	Brazil	1	2.2
	Unspecified	1	2.2
	Middle East	*2*	*4.5*
	United Arab Emirates	2	4.5
	Africa	*1*	*2.2*
	Morocco	1	2.2

Table 2 Distribution of research scope in K-12 education

Category	Characteristic	Frequency	Percentage (%)
Schooling stage	Total	45	100
	Secondary schooling	25	55.6
	Primary schooling	13	28.9
	Primary & secondary schooling	5	11.1
	K-12 schooling	2	4.4
Participants/Objects of study	Total	45	100
	Students	29	64.5
	Educators	11	24.4
	GenAI tools	5	11.1
Discipline/Research area	Total	45	100
	STEM	22	48.9
	- Science	9	20
	- Technology	7	15.6
	- Mathematics	6	13.3
	English	8	17.8
	Art	4	8.9
	History	3	6.7
	School education	3	6.7
	STEAM	2	4.4
	Religion	1	2.2
	Literacy	1	2.2
	Multiple disciplines	1	2.2

Visual analytics based on VOSviewer

VOSviewer (van Eck & Waltman, 2010) data visualization software was used for co-occurrence analysis in bibliometric analysis based on the titles and abstracts of 45 articles. Co-occurrence analysis refers to establishing connections between two keywords that appear simultaneously in a keyword set, and constructing a clustering graph based on distance based on the correlation status. Since repeating items within a single abstract can interfere with co-occurrence analysis results, the authors employed the "Boolean counting" method for co-occurrence analysis. "Boolean counting" focuses solely on whether an item appears in the title and abstract of an article, regardless of the times it occurs in the title and abstract of a specific article. In this method, each occurrence of an item is counted as 1, irrespective of its frequency in the title and abstract of a particular article. This means that the total count for an item will not exceed the total number of articles. We noticed that the initial setting of VOSviewer cannot treat a term's full name and abbreviation as synonyms, leading to the generation of two keyword nodes with the same meaning in the clustering graph. Therefore, a thesaurus file was created to avoid this situation, and 1355 items were ultimately retrieved. To prevent the interference of common words and ensure the distinguishability of the nodes displayed in the cluster diagram, the minimum number of item occurrences shown in the cluster diagram is set to 3, and the VOSviewer built-in

LinLog layout technology and modularity clustering method (Newman, 2004; Noack, 2009) were selected to normalize the cluster diagram. Then, the minimum cluster size was set to 10, and the results showed 4 clusters containing 123 items, as shown in Figure 2.

VOSviewer's co-occurrence analysis of titles and abstracts revealed 4 clusters. Larger node sizes indicate that the terms co-occur more often in the analyzed articles. The red cluster contains student, generative AI, tool, insight, educator, model, application, AI literacy, and ethics. This means that some studies involve educators using diverse teaching strategies to cultivate AI literacy among middle school students, such as teaching basic conceptual understanding of generative adversarial networks (GANs), while also focusing on discussing the application of GenAI in a wide range of fields and its ethical and moral implications.

The most prominent term in the blue cluster is study, followed by education, teacher, practice, artificial intelligence, teaching, perception, activity, and assessment. This shows that one of the main concerns of existing research is that educators use GenAI to evaluate student performance during teaching. Furthermore, some researchers collect data using mixed methods such as surveys, including interviews with participants to assess learning outcomes and capture their insights on using GenAI in specific subjects.

The green cluster includes ChatGPT, implication, learning, research, performance, motivation, feedback, effectiveness, LLM, challenge, etc. This means that some researchers use large language models as teaching tools for empirical research, and use

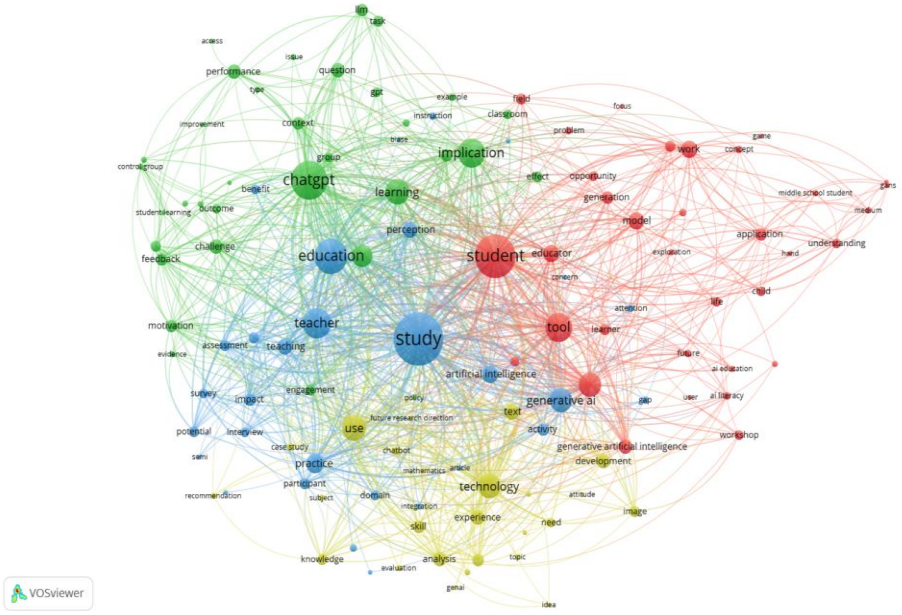

Fig. 2 Co-occurrence analysis of titles and abstracts using VOSviewer

the collected data to explore the role of feedback generated by ChatGPT in promoting teaching achievements and the impact on learners' motivation. For example, researchers set up control experiments to investigate whether the learning achievements of the experimental group have been significantly improved. However, large language models at this stage still have shortcomings and room for improvement. It is hoped that further research and innovation will be conducted on the practicality and effectiveness of large language models in teaching to meet the requirements of the education field.

The yellow cluster focuses on the overall technical aspects of GenAI, including technology, use, text, analysis, perspective, skill, knowledge, image, need, etc. This reflects the importance of developing generative artificial intelligence tools and systems designed specifically for educational scenarios. Researchers in this field face technical challenges related to algorithmic models, system architecture, and the integration of generative artificial intelligence into existing educational platforms. Developing robust and scalable generative artificial intelligence tools is crucial to ensure their practical applicability in various educational environments. Overall, the cluster analysis using VOSViewer revealed multifaceted research on GenAI in education, covering educational practices, GenAI tools, technology development, and empirical research on AI literacy.

Unique needs, directions, and practices of GenAI in K-12 education

This review revealed that scaffolding-based teaching is significant in using generative artificial intelligence in primary school. Relmasira et al. (2023) emphasized the implementation of constructivist scaffolding and promoting knowledge construction through peer discussions. Zhang et al. (2023) believed scaffolding can enhance children's participation. Kim and Park (2023) pointed out that when using GPT to generate English scripts, detailed scaffolding (such as difficulty adjustment, pronunciation assistance, and teacher adjustment) should be provided to meet personalized needs. For students who feel anxious in conversational AI, Elim (2024) suggested using the revised Bloom's taxonomy to introduce scaffolding instruction, reducing anxiety and promoting higher-order thinking. Han and Cai (2023) believed that existing AI tools are not user-friendly enough for young learners. Therefore, it is necessary to develop personalized applications based on scaffolding, constructivism, and apprenticeship models that support visual storytelling and creative expression. In the field of mathematics, Gattupalli et al. (2023) explored the use of large language models as scaffolding tools to help students overcome obstacles in problem-solving. Vartiainen and Tedre (2023) proposed providing scaffolding for children and pre-service craft teachers to promote their design and innovative thinking by gradually guiding and externalizing psychological images. Overall, the current demand and direction in the field is to design personalized, multi-level, and contextualized scaffolding support.

The cultivation of students' critical thinking cannot be ignored, which prompts them to question and evaluate the credibility of AI-generated text (Tirado-Olivares et al., 2023). Chiu (2024) emphasized that schools should regularly assess students' critical thinking and digital literacy as a prerequisite for the effective use of GenAI. Monteiro et al. (2024) pointed out that educators need to encourage students to think independently through personalized feedback and practical activities. In addition, incorporating gamification elements into teaching can help enhance students' learning engagement. Bachiri et al. (2023) explored how automated problem generation and gamified assessment can improve teaching effectiveness, while Han and Cai (2023) focused on the development of diverse visual storytelling tools to create more engaging learning experiences.

In teaching practice, Relmasira et al. (2023) used the design-based research (DBR) (Barab & Squire, 2004) method to improve the AI literacy of Indonesian primary school students, collecting data and conducting network analysis through multiple rounds of real classroom implementation. The study applied Teachable Machine to help students understand the principles of training data and model debugging and used DALL·E to discuss ethical issues involving deepfakes, guiding students to reflect on data quality and the impact of GenAI output. Kim and Park (2023) introduced textbooks and ChatGPT-generated scripts in the classroom of 27 9-year-old Korean EFL students and compared students' English performance in the form of reader theater. Students first performed with textbook scripts and then practiced versions generated by ChatGPT based on different English proficiency levels. The study found that GPT-generated scripts can better stimulate students' interest and enthusiasm for participation in most topics. Elim (2024) conducted a study in a primary school in Hong Kong S.A.R. Twenty-five fifth-grade students participated in a biweekly AI course based on ChatGPT, using the revised Bloom's taxonomy (Krathwohl, 2002) to analyze students' questions and reflections. The study found that "Creating" and "Evaluating" dominated cognitive activities, while "Applying" ability was significantly low, indicating that students had difficulty transferring cognitive skills from AI dialogue to other learning areas. The cognitive framework proposed in the study that applies Bloom's taxonomy to ChatGPT interactions can help enhance students' critical thinking using AI and support cognitive teaching practices in different contexts.

In secondary education, most studies focused on the cultivation of AI literacy. AI literacy refers to understanding, using, and critically evaluating AI, including mastering technical principles and identifying ethical risks and potential benefits (Putjorn & Putjorn, 2023). This ability helps young people grow up in the digital age and make wise decisions when faced with the impact of AI. Ali, DiPaola, Lee, Sindato, et al. (2021) highlighted the vulnerability of children's opinion formation and the need to cultivate critical digital literacy by demonstrating the spread of deep fakes on social networks. Classrooms should

use ChatGPT to promote student collaboration and knowledge construction and develop AI application capabilities and critical thinking (Chen et al., 2023). AI literacy enables young people to make wise and ethical decisions and use AI to promote innovation (Putjorn & Putjorn, 2023). Education should provide personalized support and an interactive environment to promote the improvement of students' ability to interact with AI tools (Ng et al., 2024). Before promoting large-scale model generation feedback, it is necessary to improve the AI literacy of teachers and students to ensure that teachers can guide students to evaluate and use information generated by AI (Meyer et al., 2024).

Critical thinking is a core competency of AI literacy. Educators cultivate students' ability to evaluate generative AI outputs, help them identify and question technical information, respond to false information and ethical challenges, and promote responsible academic behavior. Bitzenbauer (2023) introduced ChatGPT examples in middle school physics classes, effectively improved students' perception of ChatGPT, and explored the feasibility of meeting the needs of critical thinking cultivation. Studies have shown that teachers guiding students to analyze and optimize prompts critically are the key to developing reflective ability (Vartiainen et al., 2023). By interacting with AI through role-playing, students can inspire creative writing inspiration and cultivate critical reflection on ethics and authenticity in the creative process (Stornaiuolo et al., 2024). The reasonable introduction of ChatGPT can stimulate students' critical thinking and creative potential and help them show deeper thinking (ElSayary, 2024).

In addition to critical thinking, some studies also focused on students' self-regulation ability when using GenAI. Li et al. (2023) found that although flipped classrooms provide students with opportunities for autonomous learning, they also bring self-regulation challenges, such as procrastination, indicating that ChatGPT-assisted learning environments need to strengthen self-regulation mechanisms. Wu et al. (2024) proposed the CILA tool that combines ChatGPT, Apple Shortcuts, and LINE to promote learners to effectively build knowledge in the self-regulation process by providing real-time problem solutions and feedback records in a hybrid learning environment. Ng et al. (2024) showed that in order to cope with the challenges encountered by students in the self-regulation learning process due to individual learning style differences and insufficient feedback, SRLbot, as a personalized interactive tool, can provide targeted suggestions and instant feedback, stimulate learning initiative and improve self-regulation ability. In addition, some studies have pointed out that students' reasonable use of ChatGPT can promote their knowledge construction process and help them explore the subject content in depth (Chen et al., 2023). AI-assisted gamified learning has been proven to reduce students' cognitive load and promote their reflective knowledge construction process (Chen & Chang, 2024).

In secondary education practice, Kazemitabaar et al. (2023) integrated AI code generators into primary programming teaching through the online programming platform Coding

Steps to reduce cognitive burden and improve efficiency. In completing code creation and modification tasks, students use AI tools to obtain instant feedback and code generation support, enhancing their understanding and application of programming concepts, significantly improving task completion rate, programming correctness, and knowledge retention. Alneyadi and Wardat (2023) studied the impact of ChatGPT on the electromagnetic learning of eleventh-grade students in the United Arab Emirates. The experimental group used ChatGPT, and the control group received traditional teaching. After four weeks, the data showed that the experimental group had significantly higher post-test scores in knowledge mastery, application, and reasoning than the control group. Students believed that ChatGPT provided instant answers and explanations, enhanced interactivity and attractiveness, especially in explaining complex equations, provided visual assistance and solving homework, and provided empirical support for the application of AI in physics education. Stornaiuolo et al. (2024) explored how teenagers use the Character.AI platform as a catalyst and collaborative tool for creative writing in their daily writing practice. The researchers analyzed the interface and functions of the platform through a critical walkthrough methodology (Light et al., 2018). They found that students could develop narrative skills by interacting with AI characters, designing original characters, and rewriting dialogues. This practice not only supports students in developing digital writing skills, but also helps them critically understand how the platform affects the creative process and develop a critical understanding of generative AI in "fun" interactions.

In summary, the application of GenAI in K-12 education shows stage-by-stage differences and commonalities. In primary school, attention is paid to integrating scaffolding teaching support and gamification elements. Through the gradual guidance of teachers, young students are helped to understand and apply AI tools, and at the same time, they begin to develop their basic critical thinking. In the middle school stage, the emphasis is on the cultivation of critical thinking and the development of self-regulation ability, and on integrating AI technology into subject learning and creative expression. The review of these studies reveals the key directions of future educational practice. It reflects the particularity of students' cognitive development levels, learning needs, and educational goals at different learning stages, inspiring the design of future GenAI applications in educational practice.

RQ2 - What are the significant advantages or benefits of utilizing GenAI tools in K-12 education?

In this section, to comprehensively address the research questions raised, we first divided the GenAI tools used in K-12 education research into three categories to introduce them briefly. Then we conducted a thematic analysis of the selected articles, used a mixed

deductive and inductive approach to code the advantages of GenAI in basic education in the articles, and described the themes and sub-themes obtained.

Classification of GenAI tools

Based on the frequency statistics of using GenAI tools in research, 30 studies involved large language tools including ChatGPT, 9 studies involved image generation tools including Midjourney, 4 studies involved both GenAI language tools and image generation tools, and 2 studies involved both image generation tools and audio generation tools, as shown in Table 3. Given the possibility of using multiple GenAI tools in a study, 45 studies were recorded using GenAI tools a total of 62 times, of which large language models were counted 41 times, image generation models 18 times, and audio generation models 3 times.

Coding and thematic analysis of advantages

We coded the content of the selected publications and systematically identified codes related to the advantages of GenAI utilization in K-12 education. These codes were then organized and grouped into sub-themes, which were further consolidated into three main themes: Learning outcomes enhancement, Teaching support, and Learning experience enhancement (see Table 4).

Learning outcomes enhancement: Learning outcomes enhancement refers to the use of specific teaching strategies and methods by students to enhance their academic performance and learning skills through teaching tools. The sub-themes under this theme include Academic performance, Cognitive development, Critical thinking, and Knowledge construction. In terms of promoting student learning outcomes, students improved their grades in electromagnetics after using ChatGPT (Alneyadi & Wardat, 2023); The application of OpenAI Codex can improve the completion rate of programming tasks for students and reduce the occurrence of syntax errors (Kazemitabaar et al., 2023); The combination of Digital game-based learning (DGBL) and GPT by educators can enhance students' academic performance in physics (Chen & Chang, 2024); The SRLbot developed by researchers based on GPT-3 can effectively reduce students' learning anxiety and improve their performance in science (Ng et al., 2024); And researchers have shown that large language models can help students enhance the professionalism of expression in English writing (ElSayary 2024; Meyer et al., 2024).

Table 3 Classification of GenAI tools

GenAI tools involved	Frequency	Percentage (%)	Examples of GenAI tools
GLT	30	66.7	ChatGPT, GPT-4
IGT	9	20	Midjourney, DALL·E
GLT & IGT	4	8.9	-
IGT & AGT	2	4.4	A.I. DUET, Melody Mixer

Note. GLT: GenAI Language Tools; IGT: Image Generation Tools; AGT: Audio Generation Tools

Table 4 The advantages of utilizing GenAI tools in K-12 education

Inductive themes	Sub-themes	f	Sample study
Learning outcomes enhancement	Academic performance	8	Alneyadi & Wardat, 2023
	Cognitive development	8	Chen & Chang, 2024
	Critical thinking	5	Relmasira et al., 2023
	Knowledge construction	5	Wu et al., 2024
Teaching support	Personalized learning	11	Kim & Park, 2023
	Teaching performance enhancement	9	Tirado-Olivares et al., 2023
	Teacher work efficiency optimization	8	Zheng & Tse, 2023
	Teaching bias minimization	1	A. V. Y. Lee et al., 2023
Learning experience enhancement	Motivation enhancement	12	Chen & Chang, 2024
	Creativity development	9	Stornaiuolo et al., 2024
	Timely feedback	8	Monteiro et al., 2024
	Engagement enhancement	7	Kazemitabaar et al., 2023
	High-quality content generation	5	Gattupalli et al., 2023
	Perception and interest	5	Zhang et al., 2023
	AI literacy enhancement	4	Putjorn & Putjorn, 2023
	Collaboration facilitation	4	Chen et al., 2023
	Self-regulation	3	Ng et al., 2024
	Stress reduction	2	Kazemitabaar et al., 2023
	Information search effectiveness enhancement	1	Chen et al., 2023

In terms of Cognitive development, ChatGPT under regulation can reduce students' cognitive load (Chen & Chang, 2024) and promote the development of their cognitive abilities (Alneyadi & Wardat, 2023; ElSayary 2024; Wu et al., 2024). Applying GenAI technology to teaching AI-related knowledge can fully promote students' understanding of artificial intelligence concepts (Ali, DiPaola, & Breazeal, 2021; Putjorn & Putjorn, 2023; Williams et al., 2024). Text-to-image models can help students with special education needs better express their ideas and develop conceptual skills (Liu et al., 2023).

Multiple research results showed that large language models can provide students with inspiration and ideas, and promote collaboration to support knowledge construction (Chen et al., 2023; A. V. Y. Lee et al., 2023; Relmasira et al., 2023; Wu et al., 2024). Elim (2024) found that students will demonstrate higher-order thinking skills when using ChatGPT and think critically about the generated content. Students used Teachable Machine to train machine learning models to classify images, analyze results, and troubleshoot problems, which promoted their critical thinking abilities (Relmasira et al., 2023). Stornaiuolo et al.'s (2024) study used the Character.AI platform to encourage critical viewing of generated content and establish moral responsibility in adolescents' writing processes.

Teaching support: Teaching support refers to providing teaching resources and services to educators and learners, aiming to optimize the teaching effect of educators, enhance learners' learning abilities, and provide a comprehensive and adaptable learning

environment. Sub-themes under this theme include Personalized learning, Teacher work efficiency optimization, Teaching performance enhancement, and Teaching bias minimization. Personalized teaching is evident in a broad array of aspects. ChatGPT's powerful semantic understanding ability enables it to make customized distinctions based on students' knowledge levels and interests and can provide constructive guidance to students with different learning progress and levels (Alneyadi & Wardat, 2023; Kim & Park, 2023; Meyer et al., 2024), and flexibly adjust teaching content and strategies (ElSayary, 2024; Jauhiainen & Guerra, 2023; Monteiro et al., 2024; Ng et al., 2024). For teachers, large language models can assist English teachers in reading-related course design to save their time (J. H. Lee et al., 2023). It can also be used to generate feedback and suggestions on students' assignments, which significantly reduces teachers' workload (Meyer et al., 2024) and improves teacher productivity (Bekeš & Galzina, 2023; Chiu, 2024; ElSayary, 2024; Fassbender, 2024; Zheng & Tse, 2023).

In terms of Teaching performance enhancement, the application of text-to-image models can support teachers in optimizing the existing STEAM teaching framework (Relmasira et al., 2023) and inspire Finnish K-12 craft teachers to update the assessment and development strategies of iterative skills (Vartiainen & Tedre, 2023). ChatGPT has the potential to assist teachers in automated tutoring (Wang & Demszky, 2023; Li et al., 2023); participate in and optimize the development of teaching materials (Küchemann et al., 2023; Monteiro et al., 2024); help improve historical thinking skills in history teaching (Tirado-Olivares et al., 2023); cultivate teachers' interdisciplinary teaching capabilities (Chiu, 2024); provide inspiration and assist teachers in building high-quality curriculum planning (ElSayary, 2024). Moreover, a study suggested that ChatGPT can minimize teachers' inherent biases towards teaching content and direction, and provide an objective and fair teaching environment (A. V. Y. Lee et al., 2023).

Learning experience enhancement: In this review, Learning experience enhancement involves using GenAI and teaching strategies to improve students' motivation, engagement, and satisfaction, making the learning process more engaging, effective, and meeting the diverse needs of students. Sub-themes under this theme include Motivation enhancement, Creativity development, Timely feedback, Engagement enhancement, High-quality content generation, Perception and interest, AI literacy enhancement, Collaboration facilitation, Self-regulation, Stress reduction, and Information search effectiveness enhancement. The study by J. H. Lee et al. (2023) found that the foreign language enjoyment level and interests of students in the experimental group were significantly improved after the experiment. Bio Sketchbook developed by Zhang et al. (2023) can generate unique style outlines based on captured pictures of plants, which has been confirmed to stimulate children's interest in exploring nature. The feedback generated by ChatGPT can enhance self-confidence (Zheng & Tse, 2023) and promote students' desire

to explore and positive perceptions (Meyer et al., 2024). Studies have found that large language models can enhance intrinsic motivation in learning English (Wu et al., 2024), mathematics (Li et al., 2023), and physics (Chen & Chang, 2024; Liang et al., 2023) and promote students' active learning of scientific knowledge (Ng et al., 2024). Bachiri et al. (2023) found that compared with traditional methods, students' motivation was significantly improved through gamified assessment (Bachiri et al., 2023) and formative feedback (Zheng & Tse, 2023) supported by GenAI technology. Wu et al. (2024) found that the experimental group's cognitive, behavioral, and emotional engagement significantly improved. Text-to-image models can expand students' participation and narrow the gap between students and those who lack artistic ability and interest (J. H. Lee et al., 2023). OpenAI Codex (Kazemitabaar et al., 2023) can also increase the participation of beginners in programming.

One of the significant advantages presented in the reviewed articles is the effectiveness of promoting creativity in learners. This implies that learners have the flexibility to use their creativity to brainstorm and solve problems. Dream Studio significantly improves the performance of creative projects in terms of divergent thinking and productivity by creating numerous examples (U. Lee, Han, et al., 2023). The code provided by OpenAI Codex allows learners to draw inferences and significantly improve their creative performance (Kazemitabaar et al., 2023). Character.AI enables users to continue, modify, or create personalized stories. This collaborative method allows users to reconstruct and develop stories from multiple perspectives, thereby expanding the boundaries of creation (Stornaiuolo et al., 2024). Similarly, Midjourney provides a new medium to foster creativity for expressing students' design ideas (Vartiainen & Tedre, 2023). In history learning, participants found that the texts generated by ChatGPT were more cohesive in ideas, had highly readable structures, and had more formal wording (Tirado-Olivares et al., 2023). Participants in multiple studies have reported that ChatGPT can provide concise and explicit content in real-time to improve students' learning efficiency (Chen et al., 2023; ElSayary, 2024; Monteiro et al., 2024).

The existing studies involving K-12 students have mentioned that GenAI technology can improve users' AI literacy (Ali, DiPaola, Lee, Sindato, et al., 2021; Chen et al., 2023; Putjorn & Putjorn, 2023; Relmasira et al., 2023). GenAI tools can create supportive teaching environments to promote student collaboration (Chen et al., 2023; Liu et al., 2023; Relmasira et al., 2023) and between students and GenAI (Ng et al., 2024). Learners in blended learning involving GenAI can develop self-regulatory awareness by reviewing and reflecting on concepts (Li et al., 2023; Wu et al., 2024). SRLbot can encourage students to engage in metacognitive reflection by encouraging them to reflect on their learning strategies and processes (Ng et al., 2024). In addition, some learners explicitly attributed their reduced stress during programming tasks to using OpenAI Codex (Kazemitabaar et

al., 2023). Learners met psychological expectations, reduced stress, and overcame anxiety during math exercises (Zheng & Tse, 2023).

In general, GenAI in K-12 education is not just a teaching aid, but a profound force for change. Its applications cover multiple aspects, including personalized learning experiences tailored to individual differences among students, improving teaching effectiveness, cultivating basic cognitive abilities and critical thinking, and building a personalized educational environment promoting active participation and innovation. The frequent appearance of these topics in numerous studies highlights the significant impact of AI in education, pointing out areas where future research may bring the greatest benefits. With the continuous advancement of artificial intelligence technology, its application in education is expected to innovate teaching strategies further and improve learning outcomes, becoming a key research area worthy of continued exploration and development.

RQ3 - What challenges are associated with integrating GenAI in K-12 education?

Despite the numerous benefits of applying GenAI technology to K-12 education, researchers reported some challenges. Again, we carefully coded and categorized the content of the selected publications to answer the research question systematically. We identified several major themes, including GenAI technical limitations, Instructional support barriers, and Student development barriers (see Table 5).

GenAI technical limitations: The theme of technical limitations has the most significant number of sub-themes, and the most frequently reported by researchers are Unexpected or erroneous content. In addition, there are Technical requirements for GenAI, Prompts building, Data privacy, Bias, Copyright infringement, Limited cognitive ability, Operational efficiency, Information timeliness, Lack of iterative interaction, and Lack of emotion. Unexpected or erroneous content is reflected in various aspects of the user experience of different GenAI tools. For example, the generated content is higher than the knowledge level of students, which leads to their understanding difficulties (Kim & Park, 2023; J. H. Lee et al., 2023); LLMs generate unexpected content (Alneyadi & Wardat, 2023; Fassbender, 2024; Tirado-Olivares et al., 2023); LLMs generates incorrect information (Gattupalli et al., 2023; Wang & Demszky, 2023; Wu et al., 2024) even had syntax errors (Murgia et al., 2023); The code generated by OpenAI Codex lacks sufficient explanation (Kazemitabaar et al., 2023; Vartiainen & Tedre, 2023). Moreover, text-to-image models may also encounter the phenomenon of generated images not meeting student expectations (U. Lee, Han, et al., 2023; Liu et al., 2023; Vartiainen et al., 2023). The participants in the studies expressed the need for more technical aspects of GenAI tools, including improving English recognition with accents and providing multimedia such as animation (Alneyadi & Wardat, 2023), designing detailed scaffolding (Kim & Park, 2023) for students, improving the ability to interpret symbols and images in math problems

Table 5 The challenges with integrating GenAI in K-12 education

Inductive themes	Sub-themes	f	Sample study
GenAI technical limitations	Unexpected or erroneous content	24	Woo et al., 2023
	Technical requirements for GenAI	7	Alneyadi & Wardat, 2023
	Prompts building	4	Chen et al., 2023
	Data privacy	4	Fassbender, 2024
	Bias	4	Han & Cai, 2023
	Copyright infringement	3	Monteiro et al., 2024
	Limited cognitive ability	3	Kazemitabaar et al., 2023
	Operational efficiency	2	Alneyadi & Wardat, 2023
	Information timeliness	1	Chen et al., 2023
	Lack of iterative interaction	1	Liu et al., 2023
	Lack of emotion	1	Tirado-Olivares et al., 2023
Instructional support barriers	Difficulty in technological explanation	7	Vartiainen & Tedre, 2023
	Teaching barriers	5	Relmasira et al., 2023
	Insufficient popularization of AI literacy	4	Bekeš & Galzina, 2023
	Teacher assistance needs	4	Murgia et al., 2023
	Technology barriers in teaching	3	Williams et al., 2024
	Teacher development needs	1	Meyer et al., 2024
	Peer review needs for GenAI-generated content	1	A. V. Y. Lee et al., 2023
	Language and regional restrictions	1	Liu et al., 2023
Student development barriers	Dependency formation	6	Kazemitabaar et al., 2023
	Restriction of creativity	4	Liu et al., 2023
	Critical thinking suppression	4	Vartiainen et al., 2023
	Academic integrity	3	Fassbender, 2024
	Lack of skill transferability	1	Elim, 2024
	Tension and anxiety	1	Kim & Park, 2023
	Feeling of deception	1	Ali, DiPaola, Lee, Hong, et al., 2021
	Distrust in GenAI	1	Chen et al., 2023
	Distraction	1	Chen & Chang, 2024
	Lack of human interaction	1	ElSayary, 2024

(Mahon et al., 2023), applying gamification for education (Han & Cai, 2023), providing explanatory visual narratives for answers (Gattupalli et al., 2023), and editing tools for generated images (Liu et al., 2023).

The deployment of online LLMs in educational settings raises concerns about privacy and data security, as teachers and students are unaware of how commercial companies utilize inputs containing personal information (Fassbender, 2024; Han & Cai, 2023; Stornaiuolo et al., 2024; Wang & Demszky, 2023). Participants in multiple studies have argued that the output of LLMs may be biased (Abramski et al., 2023; ElSayary, 2024; Han & Cai, 2023; Vartiainen & Tedre, 2023) because it is difficult for developers to exclude all biased content in the datasets used to train LLMs. Additionally, participants were concerned about whether text-to-image models risk copyright infringement (Ali, DiPaola, Lee, Hong, et al., 2021; Monteiro et al., 2024; Vartiainen & Tedre, 2023).

The design of prompts when using GenAI tools is crucial, as it not only dictates the trajectory fundamentally and the scope of the resultant outputs but also affects the exhaustiveness and relevance of the output. A comprehensive prompt can significantly reduce misunderstandings and errors, and adapt to diverse application scenarios and needs. The absence of precise prompts by users to express their needs may lead to a degradation of the user experience (Chen et al., 2023; Kazemitabaar et al., 2023; Kim & Park, 2023; Ng et al., 2024). Participants in three studies reported that ChatGPT and OpenAI Codex had limited cognitive capabilities, therefore failing to understand user needs in specific scenarios accurately (Chen et al., 2023; Kazemitabaar et al., 2023; U. Lee, Jung, et al., 2023). In addition, the researchers reported that Stable Diffusion Online does not have an iterative interaction function like ChatGPT (Liu et al., 2023). Although the study by Tirado-Olivares et al. (2023) indicates that some participants believe the text generated by ChatGPT lacks emotional characteristics, this view is based on the findings of a single study and requires further evidence for support.

Instructional support barriers: Instructional support barriers involve the challenges and limitations educators face when integrating GenAI technology with instructional strategies. Sub-themes under this theme include Difficulty in technological explanation, Teaching barriers, Insufficient popularization of AI literacy, Teacher assistance needs, Technology barriers in teaching, Teacher development needs, Peer review needs for GenAI-generated content, and Language and regional restrictions. The most frequently reported challenge is the difficulty of explaining GenAI's technology. The GenAI model is described as a "black box" due to its complex and ingenious structure, which results in its lack of transparency, making it difficult for teachers to explain its working principles to students and further affecting students' understanding and acceptance of GenAI (Chen et al., 2023; ElSayary, 2024; Lyu et al., 2022).

The difficulties students encounter in teaching activities are also worth noting. For example, some students expressed difficulty in group cooperation (Relmasira et al., 2023), and confusion about the mechanisms of Variational Autoencoders (Lyu et al., 2022) and Gans (Ali, DiPaola, & Breazeal, 2021). EFL students need help reading and clarifying the content of English scripts generated by ChatGPT (Kim & Park, 2023), and students have difficulty understanding complex explanations in ChatGPT-generated content (Murgia et al., 2023). Participants in the teaching activities also face technical problems. It was difficult for participants to use the Teachable Machine to train data and manipulate the camera (Relmasira et al., 2023). Participants felt overwhelmed by the plethora of options in the user interface (Williams et al., 2024). Students with special educational needs have problems with Internet connection and typing (Liu et al., 2023). Teachers currently have to assist students in using GenAI appropriately and efficiently to complete teaching tasks (ElSayary, 2024; Kim & Park, 2023; Liu et al., 2023; Murgia et al., 2023). Teachers

urgently need to develop their AI literacy to maintain students' privacy (Han & Cai, 2023), help teachers better understand and utilize the latest technologies in education (Bekeš & Galzina, 2023; U. Lee, Jung, et al., 2023), promote students' critical thinking (Chiu, 2024), and identify potential biases in GenAI (Meyer et al., 2024).

Student development barriers: Partial studies have pointed out that applying GenAI to the teaching process may impact students' skill development, behavioral performance, emotion, and mental health. Sub-themes under this theme include Dependency formation, Restriction of creativity, Critical thinking suppression, Academic integrity, Lack of skill transferability, Tension and anxiety, Feeling of deception, Distrust in GenAI, Distraction, and Lack of human interaction. Researchers observed students relying on ChatGPT to complete English writing tasks (Chen et al., 2023) and asked for answers to physics questions without thinking deeply (Chen & Chang, 2024). The same situation occurred when students used OpenAI Codex to generate Python codes (Kazemitabaar et al., 2023). Additionally, three studies did not confirm that GenAI would make students dependent on it but expressed concerns about this (Liang et al., 2023; Monteiro et al., 2024; Vartiainen & Tedre, 2023). In the application of GenAI to K-12 education, the issue of academic integrity has attracted widespread attention. ElSayary (2024) pointed out that since the launch of ChatGPT, there have been multiple cases of cheating and misconduct. Monteiro et al. (2024) discussed the issue of using ChatGPT as a student reference tool in exams, noting that it is difficult for teachers to ensure that students are using it as a "tutor" rather than simply copying answers. In addition, teachers' opinions on this are also divided, with almost half believing that this is plagiarism or a similar violation, while the other half believes that it is not.

Incorrect use of GenAI tools can limit the development of students' creativity (Chen et al., 2023; Liu et al., 2023; Monteiro et al., 2024; Vartiainen & Tedre, 2023). Over-trust in GenAI tools can lead to a lack of critical thinking about the generated content (Chen et al., 2023; Han & Cai, 2023; Monteiro et al., 2024; Vartiainen et al., 2023). Elim (2024) used Bloom's taxonomy to reveal that students may lack the skills to apply AI conversations to other learning areas. Some EFL students reported that using ChatGPT to generate scripts that were higher than their English level made them feel stressed, nervous, and anxious (Kim & Park, 2023). GenAI-generated avatars are challenging to distinguish from real individuals, which may make students feel deceived (Ali, DiPaola, Lee, Hong, et al., 2021). One student said that ChatGPT, as a difficult-to-explain "black box", made them distrust the generated content (Chen et al., 2023). ElSayary's (2024) study highlighted that teachers believed that the misuse of ChatGPT could lead to a lack of human interaction among students and could not replace critical interaction and feedback in the classroom.

This qualitative coding study of GenAI applied to K-12 education identified several key challenges, including technical limitations, teaching barriers, and impacts on student

development. Technical issues such as erroneous content generation require improvements in the accuracy and reliability of GenAI. At the same time, the need for technology suitable for teaching highlights that GenAI tools have not yet developed more features for the education field. Teaching challenges stem from the disconnect between GenAI capabilities and teachers' AI literacy levels, revealing that teachers need targeted training in the use of AI. Concerns about student development, such as dependency and creativity suppression, suggest that GenAI may undermine rather than support key educational outcomes. In addition, educators' inability to clearly explain how GenAI technology meets student needs may exacerbate students' distrust of it. These research findings advocate for a balanced integration of GenAI in K-12 education, ensuring that technological advances are accompanied by teacher support and thoughtful consideration of student needs, aiming to use GenAI tools responsibly by educational goals and ethical standards.

Discussion

Summary of evidence

From 5256 records in the selected databases, 45 studies met the inclusion criteria and were included in this systematic review. The objective is to determine the current status of research on the application of GenAI in K-12 education, the distribution of the types of GenAI tools used, and the advantages and challenges they bring. To achieve this objective, we constructed three research questions and have successfully answered them systematically.

Among the studies involving large language models, the vast majority of researchers chose to use ChatGPT or its variants such as Open Codex (Bekeš & Galzina, 2023; Kazemitabaar et al., 2023), to study its effectiveness in assisting students' academic development. Only a few studies used large language models that the researchers themselves participated in developing (Woo et al., 2023; Zheng & Tse, 2023) and large language models from other sources (Stornaiuolo et al., 2024). The possible explanations for this are the accessibility and friendliness of ChatGPT to the public after its release and the absence of other large language models that could pose a significant threat to ChatGPT from 2023 to early 2024. Several advanced multimodal large language models have been recently launched, such as GPT-4o, Claude 3 Opus, and Gemini 1.5 Pro, and it is expected that large language models will significantly broaden their application breadth in research in the future.

As for the field of text-to-image models, there are quite a few models in the selected articles, such as Midjourney (Vartiainen et al., 2023), DALL·E (Relmasira et al., 2023), Dream Studio (U. Lee, Han, et al., 2023), Stable Diffusion Online (Liu et al., 2023), Leonardo.AI (Putjorn & Putjorn, 2023). However, while the research on LLMs covers a

wide range of topics, including mathematics, physics, chemistry, English, and history, the research on text-to-image models mainly focuses on art design and cultivating students' creativity. This is not surprising, given the differences in the design, features, and functions of the two GenAI models. LLMs are mainly used to process and generate text data and to understand and respond to users' language queries. This feature of LLMs can serve as learning assistants to provide learning suggestions for K-12 students, explain complex concepts, and generate exercises based on subject range. In addition, LLMs can grade and evaluate homework for teachers, give suggestions for making questions, and assist in preparing teaching materials. In contrast, text-to-image models have the potential to play an essential role in providing visual aids, especially in the arts, which require a lot of visual explanations. It is worth noting that due to the difficulty in quantifying participants' performance and learning outcomes in the arts, common studies involving text-to-image models collect qualitative data such as participant discourse. For example, Vartiainen and Tedre (2023) collected pre-service craft teachers' and educators' perceptions on applying Midjourney to craft courses through workshops.

Regarding the advantages of using GenAI tools in K-12 education, the most frequently mentioned advantage in the selected articles is that it can enhance students' learning motivation (Chen & Chang, 2024; Li et al., 2023; Wu et al., 2024). Unlike the traditional classroom teaching model, GenAI tools represented by LLMs can provide timely personalized feedback based on student's abilities and needs without being restricted by time and space. This personalized learning support makes students feel they can control their learning progress and have autonomy in learning. On the other hand, when students use LLMs to interact, they do not need to bear the additional social pressure that may exist in the learning process. Students can avoid the discomfort caused by making mistakes when facing classmates or teachers. The above factors enhance students' enthusiasm for learning, which is conducive to promoting the healthy development of students in the learning process and improving their academic performance.

Among the 45 articles selected for this systematic review, more than half reflected that integrating GenAI tools into teaching practice may result in the content generated by GenAI not meeting expectations or giving wrong information. The reasons for this phenomenon are complex and multifaceted. LLMs generate feedback based on training data with a knowledge cutoff date. Suppose users send information to LLMs beyond the scope of their training data. In that case, they generally receive outdated information or produce hallucinations—that is, the model generates content that seems credible but is actually fictitious. General LLMs, including ChatGPT, may not fully understand the context of a specific educational environment, resulting in answers that deviate from the expected teaching goals. General LLMs may also not be able to fully provide content that exceeds or falls below the students' understanding ability based on the cognitive

development characteristics of students of different ages. Potential biases in training data pose the risk of making the output of LLMs have a specific tendency or unfair expression. Furthermore, teachers and students typically lack an understanding of the concept of prompt engineering and do not realize that interacting with GenAI is a skill that can be optimized. This cognitive gap makes it difficult for them to clearly express their needs or set appropriate constraints in practice, thus affecting the practical application of GenAI tools in educational environments.

Educators tend to describe GenAI as a "black box" to students and avoid mentioning its specific operating principles. This simplified description is mainly because the GenAI model is based on complex neural network architecture and advanced mathematical principles, making it difficult for educators to translate such abstract technical concepts into content that K-12 students can understand. Students are confused about the source of information, which makes them distrust the information provided by GenAI (Chen et al., 2023) and affect their enthusiasm for learning.

On the other hand, some studies reported concerns about students becoming dependent on GenAI (Kazemitabaar et al., 2023; Liang et al., 2023; Monteiro et al., 2024; Vartiainen & Tedre, 2023). Such concerns include the misuse of GenAI, which may hinder the development of students' learning ability (Chen et al., 2023; Kazemitabaar et al., 2023), and the reduction of the value of human interaction and guidance in the educational process (Liang et al., 2023; Monteiro et al., 2024). Given that K-12 students' metacognition is still developing, it is difficult for them to realize the impact of relying on GenAI on their learning process. Educators should design teaching tasks that require students to think independently, while helping students develop self-reflection skills so that they can recognize the impact of GenAI on their learning methods. In addition, teachers need to systematically teach students AI literacy knowledge and emphasize the value of teacher-student interaction to cultivate students to become intelligent users of AI tools rather than dependent ones.

In summary, this systematic review made the following unique contributions to the research field of GenAI applications in K-12 education. First, unlike previous reviews that focused on general AI education, this study explored explicitly the unique educational value of generative AI. Second, this review covered a critical period from January 2020 to April 2024, which was a period of rapid development and widespread application of GenAI, provided an analysis of trends in the integration of GenAI into educational practice and research, and filled the research gap in the existing literature on the development of GenAI in K-12 education during this period. Third, based on the current overview of GenAI in K-12 education, this review identified the differences and specific needs of GenAI applications at different learning stages (primary and secondary education), providing a unique perspective for integrating GenAI education for different age groups. Finally, given

that this study considered both the advantages and challenges of integrating GenAI into K-12 education, the implications section provided educators with multi-dimensional practical implications for implementing GenAI, rather than focusing only on the theoretical level. These contributions form a framework for understanding how GenAI can transform the K-12 educational environment, maximizing the potential of GenAI in promoting student learning and teacher teaching while responsibly addressing the challenges brought about by this innovative teaching environment.

Recommendations for K-12 educators using GenAI tools

Based on a comprehensive analysis of the currently selected articles, this article provides the following recommendations for K-12 educators using GenAI tools:

- To effectively integrate GenAI technology in K-12 education and improve students' overall learning performance, educators can develop comprehensive learning objectives based on the cognitive, affective, and psychomotor domains in Bloom's taxonomy. Then, select appropriate GenAI tools based on the technological pedagogical content knowledge (originally TPCK, now known as TPACK) framework (Mishra & Koehler, 2006) and design teaching activities.
- The continued development of AI literacy among K-12 educators is critical. This literacy should go beyond simple technological familiarity and develop into a comprehensive capability that includes three levels: technology operation, teaching integration, and critical evaluation. Effective AI literacy development follows a progressive path from initial exploration to ethical guidance and works best with the support of professional learning communities. In addition, educators need to cultivate differentiated AI literacy capabilities for specific subject contexts. For example, language teachers need to focus on developing the ability to evaluate the authenticity of texts, while mathematics teachers need to focus on verifying the logic of problem-solving steps. Most importantly, educators should view GenAI literacy as an extension of existing teaching technology knowledge rather than an independent new skill, which helps reduce the cognitive burden of integrating GenAI tools.
- Evaluating the effectiveness of GenAI tools is key to ensuring their added value in the teaching process. This requires educators to track students' learning outcomes, continuously collect feedback, and conduct regular teaching evaluations, enabling them to refine their teaching framework or adopt more effective GenAI tools.

Implications

While this study systematically analyzes the application status, advantages, and challenges of GenAI in K-12 education through three specific research questions, these explorations ultimately point to a more fundamental research question, namely how to effectively integrate GenAI tools into the K-12 educational environment to achieve personalized learning and teaching innovation while addressing the associated ethical issues. In the following section, we discuss the multifaceted implications of this core question based on the aforementioned analytical results.

This systematic review revealed the considerable potential of GenAI for use in K-12 education. Given that the review period corresponded to the early stages of GenAI development, researchers and educators had limited choices for GenAI tools during this period. In most studies, the data types used for GenAI input and output were limited to text and images. Therefore, it was important to acknowledge that GenAI still faced several challenges before its broader adoption in K-12 educational practice.

The selected articles in this systematic review involving text-to-image models focused on art and creative design (U. Lee, Han, et al., 2023; Liu et al., 2023; Vartiainen & Tedre, 2023; Vartiainen et al., 2023). This disciplinary limitation highlighted the urgent need to expand the application of such generative models to more diversified disciplines. For example, text-to-image models can generate personalized vocabulary cards in language learning to deepen students' understanding and memory of vocabulary. Teachers can guide students to create images or four-frame comics containing stories and ask students to describe them to support the development of students' oral and writing skills. In science education, such as biology and chemistry, delicate biological structures and chemical reaction processes can be presented through detailed images, allowing students to intuitively see the practical application of learning content and better grasp scientific principles. Therefore, it is foreseeable that such GenAI tools have great application prospects in K-12 education.

Given the dynamic nature and rapid technological iteration of generative AI, multimodal models and teaching systems based on multimodal models are the future application trends of GenAI in K-12 education. Multimodal LLMs extend the range of data inputs they can process. They can receive and analyze various inputs, including but not limited to images and audio. Students can use multimodal models to analyze graphical data from multiple disciplines, interpret visual information, and generate explanatory feedback, which helps them form a visual understanding of knowledge. Conversational practice with AI assistants that support voice input and output provides students with a way to enhance their oral skills. Specialized GenAI tools adapted to the field of education can further enhance GenAI's teaching attributes. Copilot used in the study of Bekeš and Galzina (2023) has the ability to assist teachers in generating quizzes, lesson plans, handouts, student reports, and parent

emails. Instructional systems that integrate multiple multimodal models even have the potential to expand GenAI's output to teaching materials in slide format and instructional videos with virtual portraits.

This systematic review revealed several key ethical challenges raised by GenAI in K-12 education. These findings enriched discussions on ethical frameworks and provided concrete guidance for practitioners and policymakers. Specifically, GenAI applications in K-12 education face three key ethical challenges: in terms of data privacy, commercial GenAI tools generally lack adequate student data protection measures (Fassbender, 2024); in terms of bias, these systems inevitably reflect and reinforce sociocultural biases in training data (Vartiainen & Tedre, 2023); in terms of fairness, unequal access to GenAI tools and differences in the ability to use them may exacerbate existing educational gaps.

The authors proposed five practical recommendations to address these challenges. First, educational institutions should conduct ethical assessments of GenAI tools before adopting them and identify the differential impacts of these tools on different student groups. Second, educational institutions should include AI literacy as one of the core educational guidelines to cultivate the ability of educators and students to evaluate AI content critically. Third, multiple stakeholders should participate in dialogue mechanisms to ensure that GenAI tools reflect diverse perspectives. Fourth, educational institutions should develop ethical guidelines for K-12 environments, including data protection and bias monitoring standards, to avoid ethical issues. Finally, policymakers and education leaders should develop strategies to cover as wide a range of students as possible so that they can benefit from these advanced technologies regardless of their social and economic backgrounds. Ethical considerations should be at the core of decisions about GenAI educational applications rather than an additional part, and proactively addressing these challenges is a prerequisite for ensuring that GenAI promotes educational equity and innovation.

Overall, it is necessary to reconceptualize the paradigm for applying GenAI in K-12 education. Educators should not only view it as an auxiliary tool, but also understand it as a transformative technology that can redefine the way of teaching and learning. Therefore, teacher professional development is essential for the effective integration of GenAI. Educational institutions should invest in continuous professional AI literacy training opportunities to help teachers master the use of tools, understand how to combine GenAI tools with pedagogical principles, and establish their correct ethical awareness. Furthermore, traditional assessment methods may not be sufficient to capture the comprehensive effectiveness of innovative learning promoted by GenAI. Education researchers and practitioners should develop comprehensive assessment systems to measure students' cognitive outcomes and non-cognitive skill development. These findings provided a comprehensive roadmap for education stakeholders to guide them on enhancing

rather than replacing traditional teaching methods through GenAI, creating a more personalized, inclusive, and effective educational environment for 21st-century learners.

Limitations of the study

GenAI, represented by models such as ChatGPT, has rapidly risen in 2023. Given that the review of this review does not cover a large year span and the emerging nature of GenAI (most articles were published in 2023 and 2024), this review does not provide a line chart that covers publication years and the number of articles but rather presents it in tabular form (see Table 1). Various GenAI models have recently emerged. It is expected that research on GenAI in K-12 education will continue to increase after the deadline of this review, which means that the current review cannot include more potential related research. Since the rise of GenAI has not yet been experienced for an extended period of time and GenAI has not been widely popularized in the field of education, most of the research is recent, and the scope of the study is mainly in a limited geographical context. Further evidence is required to determine the research findings' universality and long-term effectiveness in diverse student groups.

Conclusion

In this systematic review, we investigated trends in the research field, teaching themes involved, adoption of GenAI tools, research methodology, research instruments, advantages, and challenges faced. The research subjects include K-12 students, educators, and generative artificial intelligence tools. Our systematic review indicates that the United States is the most interested in this field within English-language publications, followed by Hong Kong S.A.R. and South Korea. Researchers prefer to use GenAI tools in English and mathematics subjects, with 30 out of 45 studies involving large language models, including ChatGPT and GPT-4, 11 studies involving image generation models including text graph models, and four studies involving both. In terms of research methodology, most studies adopt mixed methods, followed by quantitative and qualitative methods. Subsequently, this article conducts a thematic analysis of the selected articles using a combination of deductive and inductive approaches. In terms of benefits, the themes and sub-themes analyzed indicate that applying GenAI tools in K-12 education can improve students' academic performance, effectively develop their cognitive abilities, enhance learning motivation, and promote personalized education. However, the analysis of the challenges indicates that 24 out of 45 reports indicate that GenAI generates unexpected or erroneous content. It can be expected that this issue may be alleviated in the future with algorithm innovation. In addition, other issues that cannot be ignored include additional technical requirements, the difficulty for educators to explain technical principles to students, students' dependence on GenAI, insufficient popularization of AI literacy by educators,

and potential copyright and privacy infringement. This article also provides several suggestions for K-12 educators to help them reduce barriers to implementing teaching activities and maximize the educational potential of GenAI. Finally, given that we are currently in the early stages of the development of generative artificial intelligence, most existing empirical research has focused on the short-term impact of GenAI tools on K-12 education. Future research will incorporate more longitudinal studies to systematically evaluate the long-term and deep-level effects of GenAI in education. In addition, the research will broaden the participation group, cover higher education, and expand the case studies to a wider range of backgrounds and geographical regions, thus comprehensively evaluating the contribution of GenAI to education inclusiveness and innovation in the global education landscape.

Appendix

Table A1 An overview of research on GenAI technology in K-12 education

Research	Context	Subject	Sample size	Age and grade level	Research methodology
Bitzenbauer (2023)	High School (Germany)	Physics	53	Grade 12	Quantitative research
J. H. Lee et al. (2023)	Elementary school (South Korea)	English	121	Aged 11-12 Grade 5 & 6	Mixed research
Alneyadi & Wardat (2023)	High School (United Arab Emirates)	Electronic magnetics	122	Grade 11	Mixed research
Vartiainen et al. (2023)	Middle school (Finland)	Art & imagination	10	Aged 14-16 Grade 8 & 9	Qualitative research
A. V. Y. Lee et al. (2023)	Primary & secondary schools (Singapore)	Science	22	Not mentioned	Qualitative research
Relmasira et al. (2023)	Primary school (Indonesia)	STEAM	55	Grade 5	Qualitative research
Chen et al. (2023)	High School (South America)	Religion	10	Not mentioned	Qualitative research
Lyu et al. (2022)	High School (USA)	Variational Autoencoders (VAE, artificial intelligence)	22	Not mentioned	Mixed research
Zhang et al. (2023)	Primary school (China)	Awareness of biodiversity	19	Aged 6-10	Mixed research
Wu et al. (2024)	High School (Taiwan, China)	Mathematics	70	Not mentioned	Quantitative research
U. Lee, Han, et al. (2023)	Elementary school (South Korea)	Art (STEAM)	46	Aged 10-11	Mixed research
Kazemitabaar et al. (2023)	Secondary school (Canada)	Programming	69	Aged 10-17	Mixed research
Woo et al. (2023)	Secondary school (Hong Kong S.A.R.)	English writing	4	Aged 15-17 Grade 7-11	Qualitative research
Ali, DiPaola, & Breazeal (2021)	Middle school (USA)	Generative Adversarial Networks (STEM)	72	Grade 5-11	Mixed research
Kim & Park (2023)	Primary school (South Korea)	English	27	Aged 9 Grade 4	Mixed research
Jauhiainen & Guerra (2023)	Primary school (Uruguay)	History	110	Aged 8-14 Grade 4-6	Quantitative research
Chen & Chang (2024)	Middle school (Taiwan, China)	Physics	202	Grade 7	Mixed research

Ng et al. (2024)	Secondary school (Hong Kong S.A.R.)	Science	74	Grade 10	Mixed research
Meyer et al. (2024)	High School (Germany)	English writing	459	Grade 10	Quantitative research
Bachiri et al. (2023)	Primary & secondary school (Morocco)	School education	100	Grade 5	Mixed research
Putjorn & Putjorn (2023)	High school (Thailand)	Generative AI	15	Aged 16-18	Mixed research
Ali, DiPaola, Lee, Sindato, et al. (2021)	Middle & high school (USA)	Generative AI	38	Aged 10-15	Mixed research
Murgia et al. (2023)	Primary school (Italy)	History	47	Aged 9-10 Grade 4	Mixed research
Stornaiuolo et al. (2024)	High school (USA)	English writing	2	Grade 12	Qualitative research
Williams et al. (2024)	Primary school (USA)	Art & imagination	8	Aged 9-11	Mixed research
Ali, DiPaola, Lee, Hong, et al. (2021)	Middle school (USA)	Generative AI	72	Grade 5-9	Mixed research
Elim (2024)	Primary school (Hong Kong S.A.R.)	English writing	25	Grade 5	Quantitative research
Zheng & Tse (2023)	Primary school (China)	Mathematics	21	Grade 4	Quantitative research
Liu et al. (2023)	Primary & secondary school (Hong Kong S.A.R.)	Art & Generative AI	12	Not mentioned	Mixed research
Liang et al. (2023)	Secondary school (Hong Kong S.A.R.)	Physics	N/A	N/A	Qualitative research
Wang & Demszky (2023)	Primary school (USA)	Mathematics	N/A	N/A	Quantitative research
Mahon et al. (2023)	High school (Ireland)	Computer science	N/A	N/A	Quantitative research
Li et al. (2023)	Middle school (Taiwan, China)	Mathematics	N/A	N/A	Quantitative research
Küchemann et al. (2023)	High school (Germany)	Physics	26	Aged 23	Quantitative research
Han & Cai (2023)	Primary school (USA)	Visual storytelling, literacy	9	Not mentioned	Qualitative research
Gattupalli et al. (2023)	K-12 (USA)	Mathematics	33	Not mentioned	Mixed research

Bekeš & Galzina (2023)	Middle school (Croatia)	Mathematics, Biology, History	34	Not mentioned	Quantitative research
U. Lee, Jung, et al. (2023)	Middle & high school (South Korea)	English	14	Not mentioned	Mixed research
Tirado-Olivares et al. (2023)	Primary school (Spain)	History	103	Not mentioned	Mixed research
Chiu (2024)	Primary & secondary school (Hong Kong S.A.R.)	Learning, teaching, assessment, administration	88	Not mentioned	Qualitative research
Vartiainen & Tedre (2023)	Primary & secondary school (Finland)	Craft education	15	Not mentioned	Qualitative research
ElSayary (2024)	Middle & high school (United Arab Emirates)	School education	40	Aged 25-54	Mixed research
Monteiro et al. (2024)	K-12 (Brazil)	Science	More than 400	Not mentioned	Mixed research
Fassbender (2024)	Secondary school (USA)	English	2	Not mentioned	Qualitative research
Abramski et al. (2023)	Secondary school (Italy)	Mathematics (STEM)	N/A	N/A	Quantitative research

Table A2 GenAI tools in K-12 studies

Research	Context / Subject	GenAI language tools	Image generation tools	Audio generation tools
Bitzenbauer (2023)	High School (German) / Physics	ChatGPT	-	-
J. H. Lee et al. (2023)	Elementary school (South Korea) / English	CopyAI (based on ChatGPT-3)	-	-
Alneyadi & Wardat (2023)	High School (United Arab Emirates) / Electronic magnetics	ChatGPT	-	-
Vartiainen et al. (2023)	Middle school (Finland) / Art & imagination	-	Midjourney	-
A. V. Y. Lee et al. (2023)	Primary and secondary schools (Singapore) / Science	ChatGPT	-	-
Relmasira et al. (2023)	Primary school (Indonesia) / STEAM	-	DALL·E	-
Chen et al. (2023)	High School (South America) / Religion	ChatGPT	-	-
Lyu et al. (2022)	High School (USA) / VAE (artificial intelligence)	-	Sketch RNN	Melody Mixer, Beat Blender
Zhang et al. (2023)	Primary school (China) / Awareness of biodiversity	-	Bio Sketchbook	-
Wu et al. (2024)	High School (Taiwan, China) / Mathematics	ChatGPT-based Intelligent Learning Aid (CILA)	-	-
U. Lee, Han, et al. (2023)	Elementary school (South Korea) / Art (STEAM)	-	Dream Studio	-
Kazemitabaar et al. (2023)	Secondary school (Canada) / Programming	OpenAI Codex	-	-
Woo et al. (2023)	Secondary school (Hong Kong S.A.R., China) / English writing	Next Sentence Generator, Next Word Generator, Next Paragraph Generator 1, Next Paragraph Generator 2	-	-
Ali, DiPaola, Lee, Hong, et al. (2021)	Middle school (USA) / Generative Adversarial Networks (STEM)	-	Sketch RNN, AI News Anchor, This Person Does Not Exist	A.I. DUET
Kim & Park (2023)	Primary school (South Korea) / English	ChatGPT	-	-
Jauhiainen & Guerra (2023)	Primary school (Uruguay) / Social sciences	ChatGPT, ChatGPT-4	Midjourney	-
Chen & Chang (2024)	Middle school (Taiwan, China) / Physics	GPT-3.5-turbo	-	-
Ng et al. (2024)	Secondary school (Hong Kong S.A.R.) / Science	SRLbot (based on ChatGPT-3)	-	-
Meyer et al. (2024)	High School (German) / English writing	GPT-3.5-turbo	-	-
Bachiri et al. (2023)	Primary & secondary school (Morocco) / School education	A fine-tuned text2text model	-	-
Putjorn & Putjorn (2023)	High school (Thailand) / Generative AI	ChatGPT	Leonardo.AI	-
Ali, DiPaola, Lee, Sindato, et al. (2021)	Middle and high school (USA) / Generative AI	-	GANs	-
Murgia et al. (2023)	Primary school (Italy) / History	ChatGPT	-	-

Stornaiuolo et al. (2024)	High School (USA) / English writing	Character.AI	-	-
Williams et al. (2024)	Primary school (USA) / Art & imagination	-	Doodlebot (Sketch-RNN)	-
Ali, DiPaola, & Breazeal (2021)	Middle school (USA) / Generative AI	-	GANs	-
Elim (2024)	Primary school (Hong Kong S.A.R.) / English writing	ChatGPT	-	-
Zheng & Tse (2023)	Primary school (China) / Mathematics	Class Optimization Master	-	-
Liu et al. (2023)	Primary & secondary school (Hong Kong S.A.R.) / Art & Generative AI	-	Stable Diffusion Online, Canva	-
Liang et al. (2023)	Secondary school (Hong Kong S.A.R.) / Physics	ChatGPT	-	-
Wang & Demszky (2023)	Primary school (USA) / Mathematics	ChatGPT	-	-
Mahon et al. (2023)	High school (Ireland) / Computer science	ChatGPT	-	-
Li et al. (2023)	Middle school (Taiwan, China) / Mathematics	ChatGPT	-	-
Küchemann et al. (2023)	High school (Germany) / Physics	ChatGPT	-	-
Han & Cai (2023)	Primary school (USA) / Visual storytelling, literacy	ChatGPT	DALL·E	-
Gattupalli et al. (2023)	Primary & secondary school (USA) / Mathematics	GPT-4, A pre-trained BART model	-	-
Bekeš & Galzina (2023)	Middle school (Croatia) / Mathematics, Biology, History	Copilot (based on GPT-3)	-	-
U. Lee, Jung, et al. (2023)	Middle & high school (South Korea) / English	ChatGPT	-	-
Tirado-Olivares et al. (2023)	Primary school (Spain) / History	ChatGPT	-	-
Chiu (2024)	Primary & secondary school (Hong Kong S.A.R.) / Learning, teaching, assessment, administration	ChatGPT	Midjourney	-
Vartiainen & Tedre (2023)	Primary & secondary school (Finland) / Craft education	-	Midjourney	-
ElSayary (2024)	Middle & high school (United Arab Emirates) / School education	ChatGPT	-	-
Monteiro et al. (2024)	Primary & secondary school (Brazil) / Science	ChatGPT	-	-
Fassbender (2024)	Secondary school (USA) / English	ChatGPT	-	-
Abramski et al. (2023)	Secondary school (Italy) / Mathematics (STEM)	GPT-3, GPT-3.5, GPT-4	-	-

Abbreviations

AI: Artificial Intelligence; AIED: Artificial Intelligence in Education; DBR: Design-Based Research; DGBL: Digital Game-Based Learning; ERIC: Education Resources Information Centre; GANs: Generative Adversarial Networks; GenAI: Generative Artificial Intelligence; LLMs: Large Language Models; PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses; SLR: Systematic Literature Review; TPACK: Technological Pedagogical Content Knowledge; WoS: Web of Science.

Acknowledgements

The authors would like to acknowledge the reviewers for their thoughtful comments and suggestions that have greatly improved the paper. This work was supported by the Teaching Development Grant [Project No. T0283] of the Education University of Hong Kong.

Authors' contributions

TZ and YCL conceptualized the study design and methodology. TZ performed data collection and drafted the manuscript under YCL's supervision. YCL and Philip LHY contributed to the manuscript revision. All authors provided critical ideas for the study design. All authors read and approved the manuscript.

Authors' information

Tianyu Zhang is currently a PhD student at the Department of Mathematics and Information Technology, The Education University of Hong Kong, Hong Kong S.A.R., China. Contact: s1146795@s.eduhk.hk

Yiu Chi Lai is an associate professor at the Department of Mathematics and Information Technology, The Education University of Hong Kong, Hong Kong S.A.R., China. Contact: yiuchi@eduhk.hk

Philip L. H. Yu is a professor at the Department of Mathematics and Information Technology, The Education University of Hong Kong, Hong Kong S.A.R., China. Contact: plhyu@eduhk.hk

Funding

This work was supported by the Teaching Development Grant [Project No. T0283] of the Education University of Hong Kong.

Availability of data and materials

The datasets generated and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Declarations

Competing interests

The authors declare that they have no competing interests.

Author details

Department of Mathematics and Information Technology, The Education University of Hong Kong, Hong Kong S.A.R.

Received: 21 December 2024 Accepted: 29 June 2025
Published online: 1 January 2026 (Online First: 9 December 2025)

References

Aboalela, R. A. (2023). ChatGPT for generating questions and assessments based on accreditations. In *Proceedings of the 13th International Conference on Advances in Computing and Information Technology*. https://doi.org/10.48550/arXiv.2312.00047

Abramski, K., Citraro, S., Lombardi, L., Rossetti, G., & Stella, M. (2023). Cognitive network science reveals bias in GPT-3, GPT-3.5 Turbo, and GPT-4 mirroring math anxiety in High-School students. *Big Data and Cognitive Computing*, *7*(3), 124. https://doi.org/10.3390/bdcc7030124

Alam, A., & Mohanty, A. (2022). Facial analytics or virtual avatars: Competencies and design considerations for student-teacher interaction in AI-powered online education for effective classroom engagement. In R. S. Tomar et al. (Eds.), *Communication, Networks and Computing. CNC 2022. Communications in Computer and Information Science, vol 1894* (pp. 252–265). Springer, Cham. https://doi.org/10.1007/978-3-031-43145-6_21

Ali, S., DiPaola, D., & Breazeal, C. (2021). What are GANs?: Introducing Generative Adversarial Networks to middle school students. *Proceedings of the AAAI Conference on Artificial Intelligence*, *35*(17), 15472–15479. https://doi.org/10.1609/aaai.v35i17.17821

Ali, S., DiPaola, D., Lee, I., Hong, J., & Breazeal, C. (2021). Exploring generative models with middle school students. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (pp. 1–13). ACM. https://doi.org/10.1145/3411764.3445226

Ali, S., DiPaola, D., Lee, I., Sindato, V., Kim, G., Blumofe, R., & Breazeal, C. (2021). Children as creators, thinkers and citizens in an AI-driven future. *Computers and Education: Artificial Intelligence*, *2*(100040), 100040. https://doi.org/10.1016/j.caeai.2021.100040

Alneyadi, S., & Wardat, Y. (2023). ChatGPT: Revolutionizing student achievement in the electronic magnetism unit for eleventh-grade students in Emirates schools. *Contemporary Educational Technology*, *15*(4), ep448. https://doi.org/10.30935/cedtech/13417

Aung, Z. H., Sanium, S., Songsaksuppachok, C., Kusakunniran, W., Precharattana, M., Chuechote, S., Pongsanon, K., & Ritthipravat, P. (2022). Designing a novel teaching platform for AI: A case study in a Thai school context. *Journal of Computer Assisted Learning*, *38*(6), 1714–1729. https://doi.org/10.1111/jcal.12706

Bachiri, Y.-A., Mouncif, H., & Bouikhalene, B. (2023). Artificial Intelligence empowers gamification: Optimizing student engagement and learning outcomes in E-learning and MOOCs. *International Journal of Engineering Pedagogy*, *13*(8), 4–19. https://doi.org/10.3991/ijep.v13i8.40853

Bahroun, Z., Anane, C., Ahmed, V., & Zacca, A. (2023). Transforming education: A comprehensive review of generative artificial intelligence in educational settings through bibliometric and content analysis. *Sustainability*, *15*(17), 12983. https://doi.org/10.3390/su151712983

Baidoo-Anu, D., & Ansah, L. O. (2023). Education in the era of generative artificial intelligence (AI): Understanding the potential benefits of ChatGPT in promoting teaching and learning. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4337484

Banh, L., & Strobel, G. (2023). Generative artificial intelligence. *Electronic Markets*, *33*, 63. https://doi.org/10.1007/s12525-023-00680-1

Barab, S., & Squire, K. (2004). Design-Based Research: Putting a stake in the ground. *Journal of the Learning Sciences*, *13*(1), 1–14. https://doi.org/10.1207/s15327809jls1301_1

Baytak, A. (2023). The acceptance and diffusion of generative artificial intelligence in education: A literature review. *Current Perspectives in Educational Research*, *6*(1), 7–18. https://doi.org/10.46303/cuper.2023.2

Bearman, M., Smith, C. D., Carbone, A., Slade, S., Baik, C., Hughes-Warrington, M., & Neumann, D. L. (2012). Systematic review methodology in higher education. *Higher Education Research & Development*, *31*(5), 625–640. https://doi.org/10.1080/07294360.2012.702735

Bekeš, E. R., & Galzina, V. (2023). Exploring the pedagogical use of AI-powered chatbots educational perceptions and practices. In *Proceedings of 2023 46th MIPRO ICT and Electronics Convention* (pp. 636–641). IEEE. https://doi.org/10.23919/mipro57284.2023.10159734

Bitzenbauer, P. (2023). ChatGPT in physics education: A pilot study on easy-to-implement activities. *Contemporary Educational Technology*, *15*(3), ep430. https://doi.org/10.30935/cedtech/13176

Bozkurt, A. (2023). Unleashing the potential of generative AI, conversational agents and chatbots in educational Praxis: A systematic review and bibliometric analysis of GenAI in education. *Open Praxis*, *15*(4), 261–270. https://doi.org/10.55982/openpraxis.15.4.609

Chen, B., Zhu, X., & Díaz del Castillo H., F. (2023). Integrating generative AI in knowledge building. *Computers and Education: Artificial Intelligence*, *5*, 100184. https://doi.org/10.1016/j.caeai.2023.100184

Chen, C., & Chang, C. (2024). Effectiveness of AI-assisted game-based learning on science learning outcomes, intrinsic motivation, cognitive load, and learning behavior. *Education and Information Technologies*, *29*(14), 18621–18642. https://doi.org/10.1007/s10639-024-12553-x

Chen, X., Zou, D., Xie, H., Cheng, G., & Liu, C. (2022). Two decades of artificial intelligence in education. *Educational Technology & Society*, *25*(1), 28–47. https://www.jstor.org/stable/48647028

Chiu, T. K. F. (2024). The impact of Generative AI (GenAI) on practices, policies and research direction in education: A case of ChatGPT and Midjourney. *Interactive Learning Environments*, *32*(10), 6187–6203. https://doi.org/10.1080/10494820.2023.2253861

Elim, E. H. S. Y. (2024). Promoting cognitive skills in AI-supported learning environments: The integration of bloom's taxonomy. *Education 3-13*, 1–11. https://doi.org/10.1080/03004279.2024.2332469

ElSayary, A. (2024). An investigation of teachers' perceptions of using ChatGPT as a supporting tool for teaching and learning in the digital era. *Journal of Computer Assisted Learning*, *40*(3), 931–945. https://doi.org/10.1111/jcal.12926

Farooq, U., & Anwar, S. (2023). ChatGPT performance on standardized testing exam -- A proposed strategy for learners. *arXiv preprint arXiv:2309.14519*. https://doi.org/10.48550/arXiv.2309.14519

Fassbender, W. J. (2024). "I can almost recognize its voice": AI and its impact on ethical teacher-centaur labor. *English Teaching Practice & Critique*, *23*(1), 104–117. https://doi.org/10.1108/etpc-08-2023-0101

Fergus, S., Botha, M., & Ostovar, M. (2023). Evaluating academic answers generated using ChatGPT. *Journal of Chemical Education*, *100*(4), 1672–1675. https://doi.org/10.1021/acs.jchemed.3c00087

Gattupalli, S. S., Lee, W., Allessio, D., Crabtree, D., Arroyo, I., Woolf, B. P., & Woolf, B. (2023). Exploring pre-service teachers' perceptions of large language models-generated hints in online mathematics learning. In *Proceedings of AIED2023 Empowering Education with LLMs- the Next-Gen Interface and Content Generation* (pp. 151–162). CEUR-WS.org. https://ceur-ws.org/Vol-3487/paper10.pdf

Giray, L. (2023). Prompt engineering with ChatGPT: A guide for academic writers. *Annals of Biomedical Engineering*, *51*(12), 2629–2633. https://doi.org/10.1007/s10439-023-03272-4

Han, A., & Cai, Z. (2023). Design implications of generative AI systems for visual storytelling for young learners. In *Proceedings of the 22nd Annual ACM Interaction Design and Children Conference* (pp. 470–474). ACM. https://doi.org/10.1145/3585088.3593867

Imran, M., & Almusharraf, N. (2023). Analyzing the role of ChatGPT as a writing assistant at higher education level: A systematic review of the literature. *Contemporary Educational Technology*, *15*(4), ep464. https://doi.org/10.30935/cedtech/13605

Jauhiainen, J. S., & Guerra, A. G. (2023). Generative AI and ChatGPT in school children's education: Evidence from a school lesson. *Sustainability*, *15*(18), 14025. https://doi.org/10.3390/su151814025

Javier, D. R. C., & Moorhouse, B. L. (2024). Developing secondary school English language learners' productive and critical use of ChatGPT. *TESOL Journal*, *15*(2), e755. https://doi.org/10.1002/tesj.755

Kang, B. (2021). How the COVID-19 pandemic is reshaping the education service. In J. Lee & S. H. Han (Eds.), *The Future of Service Post-COVID-19 Pandemic, Volume 1. The ICT and Evolution of Work* (pp. 15–36). Springer, Singapore. https://doi.org/10.1007/978-981-33-4126-5_2

Kazemitabaar, M., Chow, J., Ma, C. K. T., Ericson, B. J., Weintrop, D., & Grossman, T. (2023). Studying the effect of AI code generators on supporting novice learners in introductory programming. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (pp. 1–23). ACM. https://doi.org/10.1145/3544548.3580919

Kim, K.-H., & Kim, H. G. (2023). A study on how to create interactive children's books using ChatGPT and Midjourney. *TECHART Journal of Arts and Imaging Science*, *10*(2), 39–46. https://doi.org/10.15323/techart.2023.6.10.2.39

Kim, S., & Park, S. H. (2023). Young Korean EFL learners' perception of role-playing scripts: ChatGPT vs. Textbooks. *Korean Journal of English Language and Linguistics*, *23*, 1136–1153. https://doi.org/10.15738/kjell.23..202312.1136

Kostka, I., & Toncelli, R. (2023). Exploring applications of ChatGPT to English Language Teaching: Opportunities, challenges, and recommendations. *TESL-EJ*, *27*(3). https://doi.org/10.55593/ej.27107int

Krathwohl, D. R. (2002). A revision of Bloom's Taxonomy: An Overview. *Theory Into Practice*, *41*(4), 212–218. https://doi.org/10.1207/s15430421tip4104_2

Küchemann, S., Steinert, S., Revenga, N., Schweinberger, M., Dinc, Y., Avila, K. E., & Kuhn, J. (2023). Can ChatGPT support prospective teachers in physics task development? *Physical Review Physics Education Research*, *19*(2), 020128. https://doi.org/10.1103/physrevphyseducres.19.020128

Lee, A. V. Y., Tan, S. C., & Teo, C. L. (2023). Designs and practices using generative AI for sustainable student discourse and knowledge creation. *Smart Learning Environments*, *10*, 9. https://doi.org/10.1186/s40561-023-00279-1

Lee, J. H., Shin, D., & Noh, W. (2023). Artificial intelligence-based content generator technology for young English-as-a-foreign-language learners' reading enjoyment. *RELC Journal*, *54*(2), 508–516. https://doi.org/10.1177/00336882231165060

Lee, U., Han, A., Lee, J., Lee, E., Kim, J., Kim, H., & Lim, C. (2023). Prompt Aloud!: Incorporating image-generative AI into STEAM class with learning analytics using prompt data. *Education and Information Technologies*, *29*(8), 9575–9605. https://doi.org/10.1007/s10639-023-12150-4

Lee, U., Jung, H., Jeon, Y., Sohn, Y., Hwang, W., Moon, J., & Kim, H. (2023). Few-shot is enough: Exploring ChatGPT prompt engineering method for automatic question generation in English education. *Education and Information Technologies*, *29*(9), 11483–11515. https://doi.org/10.1007/s10639-023-12249-8

Leng, G., Zhang, G., Xiong, Y.-J., & Chen, J. (2024). CODP-1200: An AIGC based benchmark for assisting in child language acquisition. *Displays*, *82*(102627), 102627. https://doi.org/10.1016/j.displa.2023.102627

Li, P.-H., Lee, H.-Y., Cheng, Y.-P., Starčič, A. I., & Huang, Y.-M. (2023). Solving the self-regulated learning problem: Exploring the performance of ChatGPT in mathematics. In Y. M. Huang & T. Rocha (Eds.), *Innovative Technologies and Learning. ICITL 2023. Lecture Notes in Computer Science, vol 14099* (pp. 77–86). Springer, Cham. https://doi.org/10.1007/978-3-031-40113-8_8

Liang, Y., Zou, D., Xie, H., & Wang, F. L. (2023). Exploring the potential of using ChatGPT in physics education. *Smart Learning Environments*, *10*, 52. https://doi.org/10.1186/s40561-023-00273-7

Light, B., Burgess, J., & Duguay, S. (2018). The walkthrough method: An approach to the study of apps. *New Media & Society*, *20*(3), 881–900. https://doi.org/10.1177/1461444816675438

Liu, J., Kong, X., Xia, F., Bai, X., Wang, L., Qing, Q., & Lee, I. (2018). Artificial intelligence in the 21st century. *IEEE Access: Practical Innovations, Open Solutions*, *6*, 34403–34421. https://doi.org/10.1109/access.2018.2819688

Liu, X., Lau, N., Chuin, A., Leung, W. K. R., Ho, A. H. S., Das, M., Liu, M., & Kwok, C. L. (2023). Understanding students' perspectives, practices, and challenges of designing with AI in special schools. In *Proceedings of the Eleventh International Symposium of Chinese CHI* (pp. 197–209). ACM. https://doi.org/10.1145/3629606.3629625

Luo, W., He, H., Liu, J., Berson, I. R., Berson, M. J., Zhou, Y., & Li, H. (2023). Aladdin's Genie or Pandora's Box for early childhood education? Experts chat on the roles, challenges, and developments of ChatGPT. *Early Education and Development*, *35*(1), 96–113. https://doi.org/10.1080/10409289.2023.2214181

Lyu, Z., Ali, S., & Breazeal, C. (2022). Introducing Variational Autoencoders to high school students. *Proceedings of the AAAI Conference on Artificial Intelligence. AAAI Conference on Artificial Intelligence*, *36*(11), 12801–12809. https://doi.org/10.1609/aaai.v36i11.21559

Mahon, J., Mac Namee, B., & Becker, B. A. (2023). No more pencils no more books: Capabilities of generative AI on Irish and UK computer science school leaving examinations. In *Proceedings of the 2023 Conference on United Kingdom & Ireland Computing Education Research* (pp. 1–7). ACM. https://doi.org/10.1145/3610969.3610982

Meyer, J., Jansen, T., Schiller, R., Liebenow, L. W., Steinbach, M., Horbach, A., & Fleckenstein, J. (2024). Using LLMs to bring evidence-based feedback into the classroom: AI-generated feedback increases secondary students' text revision, motivation, and positive emotions. *Computers and Education: Artificial Intelligence*, *6*, 100199. https://doi.org/10.1016/j.caeai.2023.100199

Mishra, P., & Koehler, M. J. (2006). Technological Pedagogical Content Knowledge: A framework for teacher knowledge. *Teachers College Record the Voice of Scholarship in Education*, *108*(6), 1017–1054. https://doi.org/10.1111/j.1467-9620.2006.00684.x

Monteiro, F. F., Souza, P. V. S., da Silva, M. C., Maia, J. R., da Silva, W. F., & Girardi, D. (2024). ChatGPT in Brazilian K-12 science education. *Frontiers in Education*, *9*. https://doi.org/10.3389/feduc.2024.1321547

Murgia, E., Pera, M. S., Landoni, M., & Huibers, T. (2023). Children on ChatGPT readability in an educational context: Myth or opportunity? In *Adjunct Proceedings of the 31st ACM Conference on User Modeling, Adaptation and Personalization* (pp. 311–316). ACM. https://doi.org/10.1145/3563359.3596996

Newman, M. E. J. (2004). Fast algorithm for detecting community structure in networks. *Physical Review E, Statistical, Nonlinear, and Soft Matter Physics*, *69*(6), 066133. https://doi.org/10.1103/physreve.69.066133

Ng, D. T. K., Tan, C. W., & Leung, J. K. L. (2024). Empowering student self-regulated learning and science education through ChatGPT: A pioneering pilot study. *British Journal of Educational Technology*, *55*(4), 1328–1353. https://doi.org/10.1111/bjet.13454

Noack, A. (2009). Modularity clustering is force-directed layout. *Physical Review E, Statistical, Nonlinear, and Soft Matter Physics*, *79*(2), 026102. https://doi.org/10.1103/physreve.79.026102

Page, M. J., Moher, D., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … McKenzie, J. E. (2021). PRISMA 2020 explanation and elaboration: Updated guidance and exemplars for reporting systematic reviews. *BMJ (Clinical Research Ed.)*, *372*, n160. https://doi.org/10.1136/bmj.n160

Pantelimon, F.-V., Bologa, R., Toma, A., & Posedaru, B.-S. (2021). The evolution of AI-driven educational systems during the COVID-19 pandemic. *Sustainability*, *13*(23), 13501. https://doi.org/10.3390/su132313501

Puryear, B., & Sprint, G. (2022). Github copilot in the classroom: Learning to code with AI assistance. *Journal of Computing Sciences in Colleges*, *38*(1), 37–47. https://www.ccsc.org/publications/journals/NW2022.pdf#page=37

Putjorn, T., & Putiorn, P. (2023). Augmented imagination: Exploring generative AI from the perspectives of young learners. In *Proceedings of 2023 15th International Conference on Information Technology and Electrical Engineering* (pp. 353–358). IEEE. https://doi.org/10.1109/icitee59582.2023.10317680

Qin, C., Zhang, A., Zhang, Z., Chen, J., Yasunaga, M., & Yang, D. (2023). Is ChatGPT a general-purpose natural language processing task solver? In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing* (pp. 1339–1384). Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.emnlp-main.85

Rakap, S. (2023). Chatting with GPT: Enhancing individualized education program goal development for novice special education teachers. *Journal of Special Education Technology*, *39*(3), 339–348. https://doi.org/10.1177/01626434231211295

Relmasira, S. C., Lai, Y. C., & Donaldson, J. P. (2023). Fostering AI literacy in elementary Science, Technology, Engineering, Art, and Mathematics (STEAM) education in the age of generative AI. *Sustainability*, *15*(18), 13595. https://doi.org/10.3390/su151813595

Rizvi, S., Waite, J., & Sentance, S. (2023). Artificial Intelligence teaching and learning in K-12 from 2019 to 2022: A systematic literature review. *Computers and Education: Artificial Intelligence*, *4*, 100145. https://doi.org/10.1016/j.caeai.2023.100145

Roll, I., & Wylie, R. (2016). Evolution and revolution in artificial intelligence in education. *International Journal of Artificial Intelligence in Education*, *26*(2), 582–599. https://doi.org/10.1007/s40593-016-0110-3

Rüdian, S., & Pinkwart, N. (2023). Auto-generated language learning online courses using generative AI models like ChatGPT. In *21. Fachtagung Bildungstechnologien (DELFI)* (pp. 65–76). Gesellschaft für Informatik e.V. https://doi.org/10.18420/DELFI2023-14

Ruskov, M. (2023). Grimm in wonderland: Prompt engineering with Midjourney to illustrate fairytales. In *Proceedings of 19th Conference on Information and Research science Connecting to Digital and Library Science* (pp. 180–191). CEUR WS. https://doi.org/10.48550/arXiv.2302.08961

Stornaiuolo, A., Higgs, J., Jawale, O., & Martin, R. M. (2024). Digital writing with AI platforms: The role of fun with/in generative AI. *English Teaching Practice & Critique*, *23*(1), 83–103. https://doi.org/10.1108/etpc-08-2023-0103

Su, J., & Yang, W. (2023). Powerful or mediocre? Kindergarten teachers' perspectives on using ChatGPT in early childhood education. *Interactive Learning Environments*, *32*(10), 6496–6508. https://doi.org/10.1080/10494820.2023.2266490

Tirado-Olivares, S., Navío-Inglés, M., O'Connor-Jiménez, P., & Cózar-Gutiérrez, R. (2023). From human to machine: Investigating the effectiveness of the conversational AI ChatGPT in historical thinking. *Education Sciences*, *13*(8), 803. https://doi.org/10.3390/educsci13080803

van den Berg, G., & du Plessis, E. (2023). ChatGPT and generative AI: Possibilities for its contribution to lesson planning, critical thinking and openness in teacher education. *Education Sciences*, *13*(10), 998. https://doi.org/10.3390/educsci13100998

van Eck, N. J., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*, *84*(2), 523–538. https://doi.org/10.1007/s11192-009-0146-3

VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist*, *46*(4), 197–221. https://doi.org/10.1080/00461520.2011.611369

Vartiainen, H., & Tedre, M. (2023). Using artificial intelligence in craft education: Crafting with text-to-image generative models. *Digital Creativity*, *34*(1), 1–21. https://doi.org/10.1080/14626268.2023.2174557

Vartiainen, H., Tedre, M., & Jormanainen, I. (2023). Co-creating digital art with generative AI in K-9 education: Socio-material insights. *International Journal of Education through Art*, *19*(3), 405–423. https://doi.org/10.1386/eta_00143_1

Vassiliou, G., Papadakis, N., & Kondylakis, H. (2023). SummaryGPT: Leveraging ChatGPT for summarizing knowledge graphs. In C. Pesquita et al. (Eds.), *The Semantic Web: ESWC 2023 Satellite Events. ESWC 2023. Lecture Notes in Computer Science, vol 13998* (pp. 164–168). Springer, Cham. https://doi.org/10.1007/978-3-031-43458-7_31

Wang, R. E., & Demszky, D. (2023). Is ChatGPT a good teacher Coach? Measuring Zero-Shot performance for scoring and providing actionable insights on classroom instruction. In *Proceedings of the 18th Workshop on Innovative Use of NLP for Building Educational Applications* (pp. 626–667). Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.bea-1.53

Wermelinger, M. (2023). Using GitHub Copilot to solve simple programming problems. In *Proceedings of the 54th ACM Technical Symposium on Computer Science Education V. 1* (pp. 172–178). ACM. https://doi.org/10.1145/3545945.3569830

Williams, R., Ali, S., Alcantara, R., Burghleh, T., Alghowinem, S., & Breazeal, C. (2024). Doodlebot: An educational robot for creativity and AI literacy. In *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*, *2* (pp. 772–780). ACM. https://doi.org/10.1145/3610977.3634950

Woo, D. J., Wang, Y., Susanto, H., & Guo, K. (2023). Understanding English as a foreign language students' idea generation strategies for creative writing with natural language generation tools. *Journal of Educational Computing Research*, *61*(7), 1464–1482. https://doi.org/10.1177/07356331231175999

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), 3–31. https://doi.org/10.1177/07356331231191125

Xie, X., Siau, K., & Nah, F. F.-H. (2020). COVID-19 pandemic – online education in the new normal and the next normal. *Journal of Information Technology Case and Application Research*, *22*(3), 175–187. https://doi.org/10.1080/15228053.2020.1824884

Zhang, C., Zhou, Z., Hu, Y., Liu, L., Wu, J., Shao, Y., Liu, J., Zhang, L., Liu, L., Chen, H., Ying, F., & Yao, C. (2023). Observe it, draw it: Scaffolding children's observations of plant biodiversity with an interactive drawing tool. In *Proceedings of the 22nd Annual ACM Interaction Design and Children Conference* (pp. 253–266). ACM. https://doi.org/10.1145/3585088.3589380

Zhang, K., & Aslan, A. B. (2021). AI technologies for education: Recent research & future directions. *Computers and Education: Artificial Intelligence*, *2*, 100025. https://doi.org/10.1016/j.caeai.2021.100025

Zheng, W., & Tse, A. W. C. (2023). The impact of generative artificial intelligence-based formative feedback on the mathematical motivation of Chinese grade 4 students: A case study. In *Proceedings of 2023 IEEE International Conference on Teaching, Assessment and Learning for Engineering* (pp. 1–8). IEEE. https://doi.org/10.1109/tale56641.2023.10398319

Zhou, X., Van Brummelen, J., & Lin, P. (2020). Designing AI learning experiences for K-12: Emerging works, future opportunities and a design framework. *arXiv preprint arXiv:2009.10228*. https://doi.org/10.48550/arXiv.2009.10228

Publisher's Note