

Participatory Educational Research (PER)
Vol.13(2), pp. 226-247, March 2026
Available online at <http://www.perjournal.com>
ISSN: 2148-6123
<http://dx.doi.org/10.17275/per.26.27.13.2>

Id: 1793257

Generative Artificial Intelligence in Education: A Systematic Literature Review on Instructional Materials Development

Asep NURYADIN*

Department of Digital Business, Universitas Pendidikan Indonesia, Bandung, Indonesia
ORCID: 0000-0001-6226-0914

Karlimah KARLIMAH

Department of Elementary Teacher Education, Universitas Pendidikan Indonesia, Bandung, Indonesia
ORCID: 0000-0002-4292-5226

Muhammad RIZKI NUGRAHA

Department of Digital Business, Universitas Pendidikan Indonesia, Bandung, Indonesia
ORCID: 0009-0009-2127-4231

Article history

Received:
29.08.2025

Received in revised form:
23.12.2025

Accepted:
24.12.2025

Key words:

Generative Artificial Intelligence (GenAI), large language models (LLMs), ChatGPT, instructional materials, higher education, K-12 education

This study presents a systematic literature review (SLR) on the use of Generative Artificial Intelligence (GenAI) in developing instructional materials across formal educational contexts. Data were retrieved from the Scopus database using a structured search strategy following PRISMA 2020 guidelines. A total of 31 studies met the inclusion criteria, with 71% conducted in higher education and 29% in K-12. Thematic categorization shows that instructional materials developed with GenAI are dominated by textual and written materials, while the most frequently used tools are large language models (LLMs)/text generation such as ChatGPT and Copilot. The findings indicate that GenAI offers strong potential to accelerate the development and diversify the formats of instructional materials, with reported benefits including time, workload, and cost efficiency, support for educators, personalization and adaptability, improved quality and range of materials, as well as positive impacts on learning processes and outcomes. However, key challenges remain, particularly related to output quality, human readiness, technical-operational constraints, ethical issues, as well as negative impacts on learning processes and outcomes. Despite several limitations, the study highlights the importance of strengthening educators' AI literacy and skills, institutional policy support, and further empirical research regarding the use of GenAI in education.

Introduction

The integration of technology into education has continuously evolved over time, including the current emergence of Generative Artificial Intelligence (GenAI). In educational contexts, GenAI can be leveraged to enhance access to information, facilitate the teaching of complex concepts, and foster personalized learning (Farrokhnia, Banihashem, Noroozi, &

* Correspondency: asep.nuryadin@upi.edu

Wals, 2023; Rahman & Watanobe, 2023). For example, ChatGPT, a GenAI application based on Large Language Models (LLMs), has the potential to be utilized across various educational domains. Its strengths lie in its advanced natural language capabilities for generating plausible responses, its self-improvement capacity, and its ability to provide personalized, real-time feedback (Farrokhnia et al., 2023). In science education, ChatGPT has been recognized for its potential to support instructional design, rubric development, and quiz creation (Cooper, 2023). In mathematics education, it can be employed to solve mathematical problems and present foundational mathematical knowledge (Wardat, Tashtoush, AlAli, & Jarrah, 2023). Similarly, in language education, ChatGPT shows promise in facilitating second language learning (Allehyani & Algamdi, 2023; Barrot, 2023).

For educators, the presence of GenAI represents a technology that can facilitate the development of instructional materials. Instructional materials serve as essential tools for designing learning experiences and realizing educational approaches in practice (Edelson et al., 2021). Given the substantial role of instructional materials, researchers and policymakers have consistently emphasized the importance of educators' active participation in designing innovative teaching resources (Prins, Bulte, & Pilot, 2018). GenAI has been shown to support the development of lesson plans in a more efficient manner (Powell & Courchesne, 2024). It can also be employed to create personalized instructional materials tailored to students' knowledge and interests (Jauhiainen & Guerra, 2023, 2024). Nevertheless, the use of GenAI must also take into account several challenges, including ensuring the quality and accuracy of instructional materials to meet educational standards (Powell & Courchesne, 2024), as well as addressing issues of bias and ethics (Godwin-Jones, 2024).

To gain an understanding of how GenAI has been utilized in the development of instructional materials, including its benefits and challenges, one possible approach is to conduct a systematic literature review. Several systematic literature reviews on the use of GenAI in education have been published in peer-reviewed journals, including those focusing on K-12 education (Marzano, 2025), vocational education (Ranuharja et al., 2025), higher education (Amofa et al., 2025; Deroncele-Acosta, Sayán-Rivera, Mendoza-López, & Norabuena-Figueroa, 2025), special education (M. Wang, Tlili, Khribi, Lo, & Huang, 2025), English as a Foreign Language (EFL) education (Liu, Sihes, & Lu, 2025), and computing education from K-12 through graduate levels (Agbo, Olivia, Oguibe, Sanusi, & Sani, 2025).

Existing systematic reviews on GenAI in education mostly focus on specific levels or disciplines, with none specifically examining the development of instructional materials across formal educational settings. Although Ranuharja et al. (2025) addressed this issue, their study was limited to vocational education. Hence, a broader review is needed to capture recent developments, support educators and stakeholders in leveraging GenAI effectively, and prepare prospective teachers with the necessary skills. Further empirical research is also required, particularly given the rapid growth of GenAI tools such as ChatGPT, Gemini, and Copilot (Al-Kadi & Ali, 2024; Leite, 2024; Nikolic et al., 2024).

Building on the above discussion, this study aims to review the literature, synthesize findings, and discuss the utilization of GenAI in the development of instructional materials, particularly within the context of formal education. For the purposes of this study, instructional materials are broadly defined to encompass textbooks, audio, video, visuals, digital media, and also assessment resources such as test items and evaluation instruments. Specifically, this study seeks to address the following research questions:

- RQ1. What types of instructional materials have been developed using GenAI in educational contexts?
- RQ2. What types of GenAI tools have been utilized to develop instructional materials?
- RQ3. What benefits are reported from the use of GenAI in developing instructional materials?
- RQ4. What challenges are reported from the use of GenAI in developing instructional materials?

By addressing these research questions, the study aims to provide insights not only into the types of instructional materials that can be developed using GenAI, but also into the types of GenAI tools that can be employed, the benefits of using GenAI for instructional materials development, and the challenges that educators and other stakeholders need to address.

Methods

To address these research questions, this study employed the systematic literature review (SLR) method. This approach applies rigorous methodological procedures in defining search and screening criteria to ensure comprehensive results and minimize the risk of bias (Labadze, Grigolia, & Machaidze, 2023). The implementation of the systematic literature review in this study followed the reporting guideline of the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement, updated in 2020 (Page et al., 2021). PRISMA was selected because it is the most widely adopted protocol for systematic literature reviews with thousands of citations (over 60,000 in Scopus as of August 2020), endorsement by nearly 200 journals and organizations, and application across multiple disciplines (Page et al., 2021).

Search process

To identify relevant literature for review, a search was conducted in the Scopus database on April 21, 2025. Scopus was selected due to its broad content coverage, credible impact indicators, and the availability of publications relevant to human-computer interaction (Aravantinos et al., 2024). The search was performed using a search string constructed with Boolean operators, as presented in Table 1. No timeframe restrictions were applied, given that peer-reviewed publications on the use of GenAI in education remain limited (Halaweh, 2023). The initial search in Scopus yielded 292 records, which were subsequently filtered based on the predetermined inclusion and exclusion criteria.

Table 1. Search Database and Search String

| Search Database | Search String |
|-----------------|---|
| Scopus | ((“ChatGPT” OR “generative artificial intelligence” OR “generative AI” OR “GenAI” OR “GAI”) AND (“teaching material” OR “teaching materials” OR “learning material” OR “learning materials” OR “instructional material” OR “instructional materials” OR “educational content” OR “educational contents” OR “lesson material” OR “lesson materials” OR “teaching resource” OR “teaching resources”)) |

Inclusion and exclusion criteria

This review process followed the predetermined inclusion and exclusion criteria to ensure transparency, achieve comprehensive results, and minimize the risk of bias. Table 2 presents these criteria along with their justifications, with particular emphasis on the role of educators in utilizing GenAI for the development of instructional materials. The established



criteria cover aspects such as study focus, educational context, type of publication, research design, accessibility, and language.

Table 2. Inclusion and Exclusion Criteria

| Inclusion | Exclusion | Justification |
|--|---|--|
| Studies focusing on the use of GenAI for the development of instructional materials in educational settings. | Studies focusing on the direct use of GenAI by students only as a learning tool, without links to instructional materials development. | To ensure alignment with the review aim, which emphasizes the development of instructional materials. |
| Studies conducted in formal educational settings (primary, secondary, or higher education) Peer-reviewed journal articles | Studies conducted in informal education, corporate training, or non-academic contexts Conference proceedings, books, book chapters, or other non-peer-reviewed sources | To focus on structured educational environments where instructional materials are systematically used To ensure methodological rigor by including only studies that have undergone formal peer review |
| Primary empirical research (qualitative, quantitative, or mixed methods) Full-text studies | Theoretical papers, literature reviews, or conceptual papers Abstract-only or inaccessible studies | To focus on original studies that contribute first-hand evidence on the topic To enable comprehensive and accurate data extraction and analysis |
| Open access sources | Non-open access resources | To promote transparency and reproducibility by allowing all included studies to be freely accessible |
| Studies written in English | Studies written in languages other than English | To ensure consistent interpretation and avoid translation bias |

These criteria subsequently served as the basis for screening and selecting publications, as illustrated in the PRISMA flow diagram (Figure 1).

Data extraction and analysis

As shown in Figure 1, the initial literature search in Scopus yielded 292 records. During the initial screening stage, 14 records were removed, consisting of 2 duplicates, 11 irrelevant records that were proceedings or conference titles rather than research articles, and 1 record excluded due to a irrelevant context, where the term GAI referred to Glucagon Autoinjector instead of Generative AI. Consequently, 278 records proceeded to the title and abstract screening stage.

Then, 229 records were excluded for not meeting the inclusion criteria, leaving 49 reports for full-text retrieval. Of these, 14 reports could not be obtained in full text. Consequently, 35 reports were successfully accessed and further evaluated at the eligibility stage. During the eligibility assessment, 4 reports were excluded based on the inclusion-exclusion criteria: 2 studies focused on the use of Generative AI by students who were not pre-service teachers, 1 lacked clarity regarding the educational context, and 1 was not written in English.

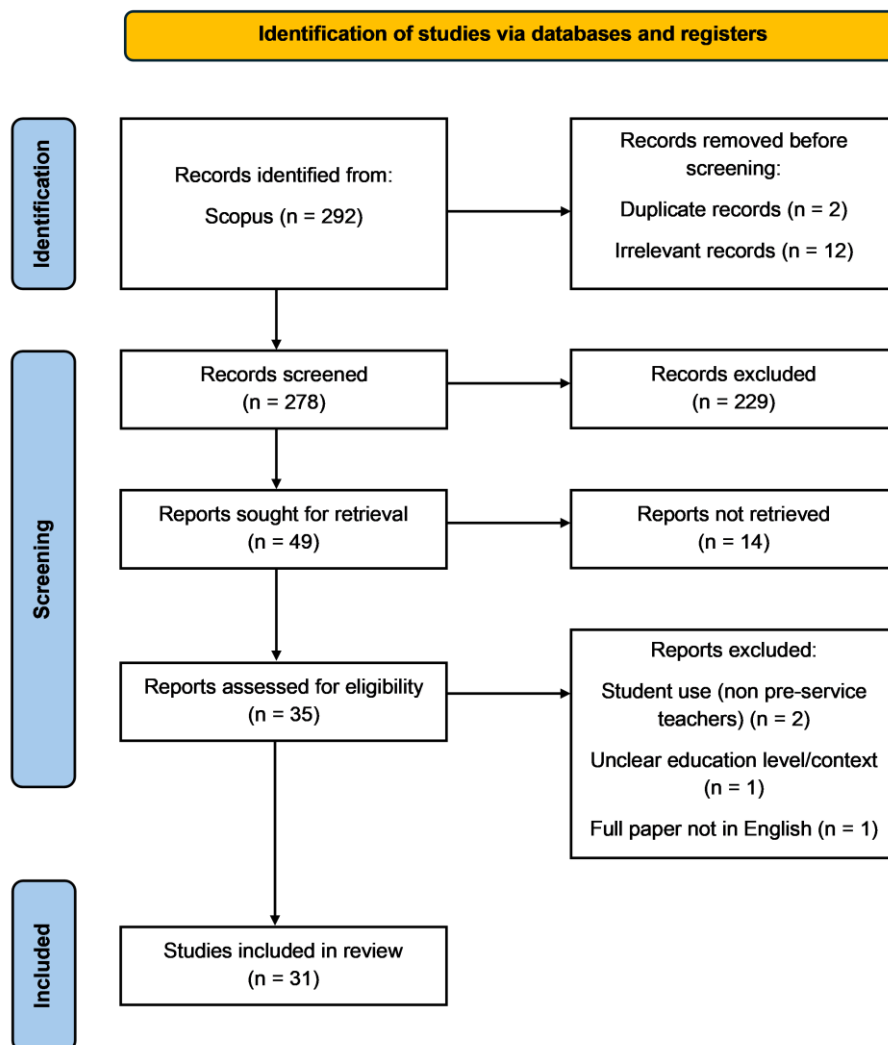


Figure 1. PRISMA 2020 Flow Diagram Illustrating the Study Selection Process (Adapted from Page et al., 2021)

Finally, 31 studies were included in the final review and thematically analyzed to identify the types of instructional materials developed, the GenAI tools used, as well as the reported benefits and challenges. In this study, a structured, theme-based approach was employed, as it allows for the development of themes and deeper insights into the field, in contrast to methods such as bibliometric reviews that primarily emphasize statistical patterns and publication trends (Shree, Singh, Paul, Hao, & Xu, 2021).

Results

This section presents the synthesized findings from the 31 studies that met the inclusion criteria. The analysis begins with a descriptive analysis of the distribution of studies by year of publication and educational level, as shown in Figure 2, to provide an overview of the context of GenAI adoption in instructional materials development. The review findings are then presented in line with the research questions (RQs), namely: (1) the types of instructional materials developed using GenAI (RQ1), (2) the types of GenAI tools employed

in instructional materials development (RQ2), (3) the reported benefits of using GenAI (RQ3), and (4) the reported challenges associated with its use (RQ4).

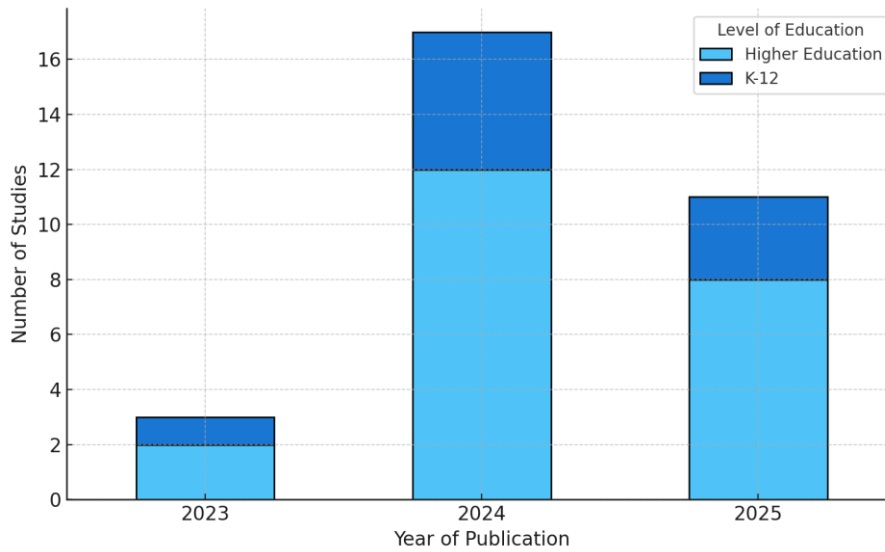


Figure 2. Distribution of Reviewed Studies by Year and Educational Level

As shown in Figure 2, only three studies met the inclusion criteria in 2023, but this number rose sharply to 17 in 2024 before slightly declining to 11 in early 2025. However, the number of studies in 2025 is likely to increase, as the data for this review were collected in April 2025. Higher education consistently dominated across years, accounting for 71% of the total 31 studies compared to 29% in K–12. This distribution suggests that research on GenAI-based instructional material development has thus far been concentrated primarily within higher education contexts.

Types of instructional materials developed using GenAI

To address the first research question, the reviewed studies were analyzed to identify the types of instructional materials developed using GenAI across various educational contexts. The analysis was conducted through thematic categorization to synthesize the diverse outputs reported in these studies. As presented in Table 3, four main categories of instructional materials were identified: (1) text and written materials, (2) assessment and evaluation, (3) multimedia, and (4) game-based and gamified materials.



Table 3. Thematic Categorization of Instructional Materials Developed Using Generative AI in the Reviewed Studies

| Category | Types of Materials | Studies (Alphabetical) |
|--|---|--|
| Textual and Written Materials | Lesson plans, syllabi, lecture texts, newsletters, vocabulary lists, conversation scripts, transcripts (monologues/dialogues), reading passages, writing samples (stories, poetry), explanatory discourse, technical definitions & coding explanations, mini-cases. | (Alkaabi & Almaamari, 2025; Byrne, 2024; Davis, 2025; Garcia-Varela, Bekerman, Nussbaum, Mendoza, & Montero, 2025; Jauhiainen & Guerra, 2023, 2024; Jin et al., 2025; Kwack & Im, 2023; Lo, 2025; Madunić & Sovulj, 2024; Mulyani, Istiaq, Shauki, Kurniati, & Arlinda, 2025; Nguyen, 2024; Noveski, Jeroncic, Velard, Kocuvan, & Gams, 2024; Pesovski, Santos, Henriques, & Trajkovik, 2024; Romaniuk & Łukasiewicz-Wieleba, 2024; Xin, 2024; Yang & Markauskaite, 2025; Zhang, 2025; Zheng & Stewart, 2024). |
| Assessment and Evaluation Materials | Educational questions, quizzes, assignments, rubrics, script concordance tests (SCTs), reframed problems, open-ended & MCQs, self-assessment, test items. | (Al Faraby, Romadhony, & Adiwijaya, 2024; Bernal, Bolender et al., 2024; Davis, 2025; Einarsson, Lund, & Jónsdóttir, 2024; Hudon, Kiepora, Pelletier, & Phan, 2024; Jauhiainen & Guerra, 2023, 2024; Jin et al., 2025; Lohr, Berges, Chugh, Kohlhase, & Müller, 2025; Pesovski et al., 2024; Zhang, 2025). |
| Multimedia Materials | Videos, audios, images/illustrations, slides, multimedia learning resources. | (Carbonell-Alcocer, Sanchez-Acedo, Benitez-Aranda, & Gertrudix, 2025; Jauhiainen & Guerra, 2023, 2024; Lo, 2025; Mulyani et al., 2025; Reed & Dodson, 2024; Romaniuk & Łukasiewicz-Wieleba, 2024; P. Wang, Jing, & Shen, 2025; Xu et al., 2025; Yang & Markauskaite, 2025; Zhang, 2025). |
| Game-based and Gamified Materials | Game-based learning environments, gamified exercises. | (Ho & Lee, 2023; Naatonis, Rusijono, Jannah, & Malahina, 2024). |

The thematic categorization presented in Table 3 indicates that text and written instructional materials are the most dominant, developed in 19 studies. Assessment and evaluation materials appeared in 12 studies, followed by multimedia in 11 studies, while game-based and gamified materials were relatively rare (2 studies). It should be noted that although the assessment and evaluation category was thematically separated, most of its products were still text-based outputs, such as multiple-choice questions, open-ended questions, and rubrics. Thus, the use of GenAI in instructional materials development remains heavily focused on the production of textual outputs rather than multimedia or game-based and gamified materials.

First, the category of text and written materials was identified in most of the studies, including the development of lesson plans (Alkaabi & Almaamari, 2025; Davis, 2025; Kwack & Im, 2023; Mulyani et al., 2025; Romaniuk & Łukasiewicz-Wieleba, 2024) and the preparation of syllabi (Madunić & Sovulj, 2024). Other types within this category included creative texts such as moral stories or poetry (Jin et al., 2025; Zheng & Stewart, 2024), and technical definitions and programming code explanations (Noveski et al., 2024).

Second, the category of assessment and evaluation materials demonstrated considerable diversity, ranging from educational questions generated by LLMs and shown to be comparable to those produced by experts (Al Faraby et al., 2024), to the development of



dynamic multiple-choice quizzes within the Learnix e-learning platform (Bernal, 2024). In medical education, GenAI was employed to design script concordance tests (SCTs) (Hudon et al., 2024). In addition, GenAI was applied for problem reframing in statistics across 17 domains (Biology, Chemistry, Economics, History, Medicine, Psychology, etc.) (Einarsson et al., 2024).

Third, the category of multimedia materials encompassed a wide range of innovations, from AI-generated instructional videos that integrated text, visuals, and bilingual explanations for language learning (Xu et al., 2025) to AI-generated patient backstory videos used to enhance clinical simulation in nursing education (Reed & Dodson, 2024). On a broader scale, GenAI was leveraged in the production of open educational resources, including interactive presentations and high-quality animated videos (Carbonell-Alcocer et al., 2025).

Fourth, the category of instructional game-based and gamified materials remains relatively limited, yet it indicates a new direction in the use of GenAI for learning. Ho and Lee (2023) integrated ChatGPT into the development of a Roblox-based learning environment for an electrical engineering course, enabling more flexible game content creation. Meanwhile, Naatonis et al. (2024) demonstrated how the integration of the ChatGPT API into a problem-based gamification learning (PBGL) system could be employed for programming practice with levels, points, and automated feedback.

The distribution of studies across each category is further illustrated in Figure 3, which highlights the dominance of text and written materials over the other types.

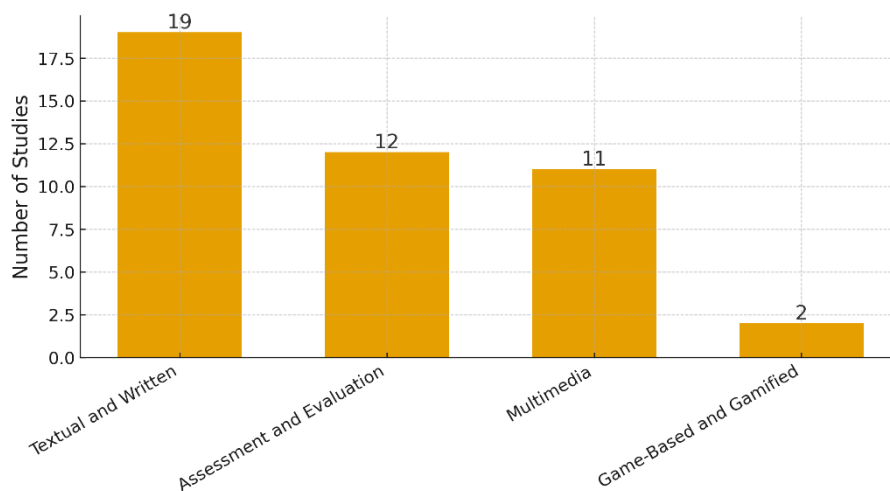


Figure 3. Distribution of Reviewed Studies by Types of Instructional Materials Developed Using GenAI

Types of GenAI tools utilized to develop instructional materials

In the development of these various types of instructional materials, a wide range of GenAI tools were employed. It is therefore necessary to analyze and categorize the types of GenAI used. This mapping is important as it provides educators with an overview of the available GenAI technologies and how each type can be leveraged to support instructional material production. A summary of the categorization of GenAI tools utilized is presented in Table 4, while their distribution is visualized in Figure 4.

Table 4. Thematic Categorization of Generative AI Tools Utilized in the Reviewed Studies

| Category | GenAI Tools | Studies (Alphabetical) |
|--|--|--|
| Large Language Models (LLMs / Text Generation) | ChatGPT (3.5, 4, 4o, Turbo); ChatGPT API; GPT-4 (LMS-integrated); LLaMA-2; Claude-3-Opus; Google Gemma-7B; Copilot; Bard (Gemini); Auto-GPT. | (Al Faraby et al., 2024; Alkaabi & Almaamari, 2025; Bernal, 2024; Byrne, 2024; Carbonell-Alcocer et al., 2025; Davis, 2025; Einarsson et al., 2024; Garcia-Varela et al., 2025; Ho & Lee, 2023; Hudson et al., 2024; Jauhiainen & Guerra, 2023, 2024; Jin et al., 2025; Kwack & Im, 2023; Lo, 2025; Lohr et al., 2025; Madunić & Sovulj, 2024; Mulyani et al., 2025; Naatonis et al., 2024; Nguyen, 2024; Noveski et al., 2024; Pesovski et al., 2024; Romaniuk & Łukasiewicz-Wieleba, 2024; Yang & Markauskaite, 2025; Zhang, 2025; Zheng & Stewart, 2024). |
| Speech and Voice Generation (Text-to-Speech) | and PlayHT; ElevenLabs; Narakeet; Resemble; Murf; Edge TTS; EaseUS. | (Carbonell-Alcocer et al., 2025; Lo, 2025; Xu et al., 2025; Yang & Markauskaite, 2025). |
| Image Generation | Microsoft Designer; Adobe Firefly; Freepik; Midjourney; Lumiere3D; Craiyon; Fotor; PIXLR; Deep Dream Generator; Bing Image. | IA; (Carbonell-Alcocer et al., 2025; Jauhiainen & Guerra, 2023, 2024; Kazanidis & Pellas, 2024; Reed & Dodson, 2024; Romaniuk & Łukasiewicz-Wieleba, 2024; Yang & Markauskaite, 2025). |
| Video Animation Generation | and Synthesia; Sudowrite; Visla; Jasper; Animaker; Lumen5; SADTalker. | (Carbonell-Alcocer et al., 2025; Kazanidis & Pellas, 2024; Xu et al., 2025). |
| Hybrid Specialized Generation Tools | / Finetune Generate; Finetune Catalog; ChatPDF; Custom GPT “AI Teacher”. | (Bolender et al., 2024; Xin, 2024; Xu et al., 2025) |

This review shows that large language models (LLMs) such as ChatGPT, Gemini, Copilot, Claude, and LLaMA are the most widely used category in the development of GenAI-based instructional materials. The flexibility of LLMs is evident across diverse educational domains. For example, these models have been used to generate dialogue transcripts that model broader communication strategies compared to textbooks (Nguyen, 2024), and to create moral stories aligned with cultural contexts in English as a Foreign Language instruction (Zheng & Stewart, 2024). LLMs have also been integrated into game-based learning, such as in the development of a Roblox-based learning environment supported by 3D modeling (Ho & Lee, 2023), and employed to generate and classify educational questions (Al Faraby et al., 2024).

Beyond LLMs, there is a clear trend toward multimodal and specialized applications. Several studies have highlighted the use of speech synthesis technologies for producing listening materials, such as PlayHT and ElevenLabs in the production of a Massive Open Online Course (Carbonell-Alcocer et al., 2025), various text-to-speech (TTS) platforms including Speechgen, Narakeet, PlayHT, Resemble, and Murf in English language learning (Lo, 2025), Microsoft Edge TTS for instructional videos (Xu et al., 2025), and EaseUS as a voice-over generator for converting listening scripts into audio in language teacher education (Yang & Markauskaite, 2025).

In addition, several studies employed image generators to enrich instructional materials. Midjourney, for instance, was used in elementary history education to produce visual illustrations that support learning content (Jauhiainen & Guerra, 2023, 2024), as well as in nursing education to create patient backstories (used in patient backstory videos) prior to



clinical simulations (Reed & Dodson, 2024). Meanwhile, in online course development, tools such as Microsoft Designer, Freepik IA, and Adobe Firefly were utilized to produce visual materials as part of the creation of open educational resources (Carbonell-Alcocer et al., 2025).

Several studies have also integrated video and animation generators to enrich learning. Synthesia, for example, was used in conjunction with Adobe Animate and MS Teams in online course production to create avatars and instructional animations (Carbonell-Alcocer et al., 2025). Xu et al. (2025) employed a combination of custom GPT “AI Teacher,” Microsoft Edge TTS, and SADtalker to generate English vocabulary instructional videos incorporating text, visuals, and bilingual explanations. In addition, the use of Visla, Jasper, Animaker, and Lumen5 for developing project-based instructional videos (Kazanidis & Pellas, 2024) further illustrates the growing adoption of GenAI for educational video content production.

In addition, several studies developed specialized or hybrid systems for specific instructional needs. For instance, Finetune Generate and Finetune Catalog were employed for creating and tagging assessment items (Bolender et al., 2024). ChatPDF was used to modify readings and design tasks (Xin, 2024), while a custom GPT “AI Teacher,” integrated with optical character recognition (OCR) and TTS, was applied to produce multimodal instructional videos (Xu et al., 2025).

The distribution of categories is further illustrated in Figure 4, which shows the frequency of utilization of different types of generative AI tools in the reviewed studies. The diagram clearly demonstrates that LLMs dominate significantly over the other categories, indicating that text-based models remain the primary choice for educators in instructional materials development. In contrast, voice synthesis, image, and video tools appear in smaller proportions but reveal a growing trend toward multimodal content production. Hybrid or specialized tools occupy a distinct position, reflecting a new direction in the use of customized applications within the learning ecosystem.

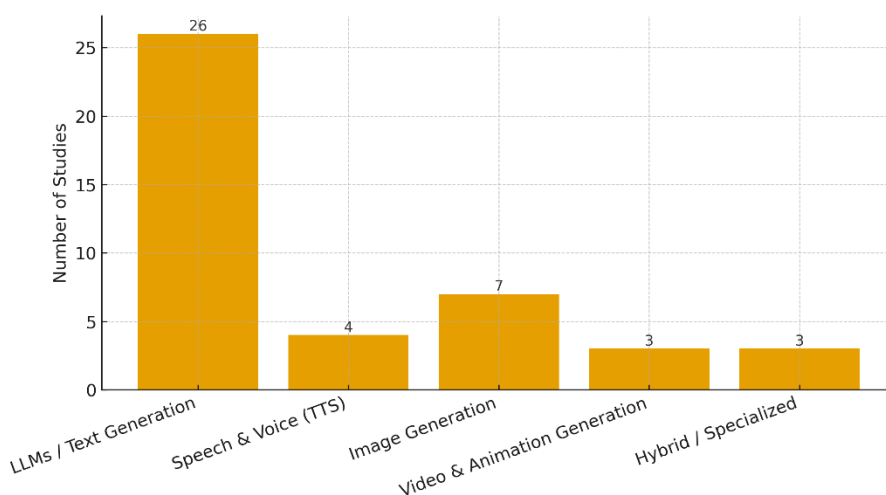


Figure 4. Distribution of Generative AI Tools Utilized in the Reviewed Studies

Reported benefits of using GenAI in instructional materials development

Efficiency in time, workload, and cost

Several studies indicated that GenAI reduces the time required and lowers the workload involved in developing instructional materials across different contexts. For instance, GenAI enabled educators to generate lesson plans more quickly (Alkaabi & Almaamari, 2025; Davis, 2025; Jin et al., 2025; Kwack & Im, 2023; Madunić & Sovulj, 2024). In language materials development, vocabulary lists and reading passages were produced more efficiently compared to non-GenAI methods (Byrne, 2024; Jin et al., 2025). In the domain of game-based learning environments, scripting and design processes became streamlined, thereby accelerating the creation of interactive content (Ho & Lee, 2023).

From the perspective of iteration, evidence suggests that usable drafts can be achieved with short prompt cycles. For instance, two prompts were sufficient to generate contextually appropriate moral dilemma stories (Zheng & Stewart, 2024), and definitions of concepts in a machine learning curriculum were produced “within seconds” with a low average number of iterations (Noveski et al., 2024).

In the area of assessment, test item development was shown to be more efficient, whether in the form of comprehension questions (Byrne, 2024), multiple-choice questions (MCQs) for self-assessment (Pesovski et al., 2024), or various formats including MCQs, True-False, and Fill-in-the-blank (Zhang, 2025). Furthermore, the development of script concordance tests was reported to be faster than if produced entirely by experts (Hudon et al., 2024). In addition, automated pipelines helped reduce the effort required for item correction compared to manual processes (Lohr et al., 2025).

Beyond efficiency in time and workload, other studies also emphasized cost efficiency when GenAI was applied in both instructional production and assessment. At the program production level, a GenAI-supported massive open online courses (MOOCs) pipeline enabled faster processes at lower costs compared to conventional workflows (Carbonell-Alcocer et al., 2025). In the context of large-scale assessment, Bolender et al. (2024) reported that a project aligning 90,000 test items in a national item bank achieved an 88% reduction in time and up to 85% cost savings compared to manual methods, while maintaining high alignment accuracy (96%) and strong expert validation (86-93%).

Support for educators

In addition to saving time, GenAI provides tangible support for educators. It sparks new ideas for designing or revising materials (Lohr et al., 2025; Yang & Markauskaite, 2025) and facilitates the search for instructional materials (Kwack & Im, 2023; Romaniuk & Łukasiewicz-Wieleba, 2024). The use of GenAI also helps educators refine explanations to make them more concrete and precise (Garcia-Varela et al., 2025). Furthermore, this technology can provide pedagogical suggestions (Xin, 2024) and strengthen reflective practice and critical engagement (Jin et al., 2025). Other forms of support include proofreading and language refinement of lesson plans (Romaniuk & Łukasiewicz-Wieleba, 2024) as well supporting problem-solving skills (Kwack & Im, 2023).

Personalization and adaptability

Overall findings indicate that GenAI facilitates the customization of learning materials and student learning pathways. In classroom contexts, educators reported GenAI's ability to personalize or customize a wide range of instructional materials to align with students' needs and/or proficiency levels (Al Faraby et al., 2024; Alkaabi & Almaamari, 2025; Jauhiainen & Guerra, 2023, 2024; Mulyani et al., 2025). Personalization was also implemented through systems such as the Learnix platform, which featured dynamic MCQs and personalized feedback (Bernal, 2024), as well as a web-based PBGL system integrated with the ChatGPT API that provided automatic feedback and adaptive learning pathways based on students' progress (Naatonis et al., 2024). Thus, personalization occurred not only at the level of content, but also through the feedback mechanisms that accompanied it.

Quality and range of instructional materials

GenAI not only enhances the quality of instructional materials but also broadens the range of formats and representations available to educators. In terms of content, AI-generated conceptual definitions in a machine learning (ML) curriculum were found to approximate expert standards with relatively few iterations (Noveski et al., 2024). From the perspective of material architecture, a retrieval-augmented generation (RAG)-based pipeline was able to produce structured semantic annotations and enrich the collection of learning objects available for instructors (Lohr et al., 2025).

Quality improvements were also observed in multimedia instructional materials. For example, in MOOCs content production, GenAI facilitated the creation of HD videos, synthetic voices, and avatars that were previously difficult to achieve at low cost (Carbonell-Alcocer et al., 2025). Another example was in a classroom project involving early childhood education students, where the use of GenAI improved the visual quality of educational videos, images, and animations produced (Kazanidis & Pellas, 2024).

In assessment, quality was enhanced through accurate alignment in large-scale projects (Bolender et al., 2024) and script concordance tests of comparable quality to those produced by experts, consistent with the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5) (Hudon et al., 2024).

Alongside these quality improvements, the range of instructional materials also expanded. From a pedagogical perspective, GenAI provided broader models of communication strategies (Nguyen, 2024), offered listening exposure across diverse varieties of English (Lo, 2025), and supported the writing of culturally appropriate stories for English as a Foreign Language (EFL) learning through systematic prompting procedures (Zheng & Stewart, 2024).

Another example can be found in nursing education, where Midjourney was used to generate images that served as the basis for creating narrative patient backstory videos (Reed & Dodson, 2024). In addition, multimodal packages emerged that combined generative text, illustrations, and tiered questions adapted to students' knowledge levels (Jauhiainen & Guerra, 2023, 2024), as well as text-image-audio combinations with voice-over to support more contextualized communicative activities (Yang & Markauskaite, 2025).

Learning processes and outcomes

The use of GenAI in developing instructional materials also influenced learning processes and outcomes. In video-based learning, groups using AI-generated instructional

videos demonstrated higher retention compared to those using traditionally recorded instructional videos (Xu et al., 2025). In primary education, materials adapted to students' proficiency levels were associated with increased enjoyment and motivation to learn, with a significant proportion reporting that they learned much or very much (Jauhiainen & Guerra, 2023, 2024). In language instruction, GenAI-assisted listening and reading tasks enhanced student engagement in class (Lo, 2025), while the GenAI-supported layered language learning approach showed improvements in reading components (Byrne, 2024).

In probability and statistics courses, reframing problem contexts added educational value and real-world relevance according to participants' evaluations (Einarsson et al., 2024). At higher cognitive levels, LLM-generated questions were found to stimulate critical thinking when designed with proper prompting (Al Faraby et al., 2024). Beyond cognitive outcomes, affective dimensions were also supported; for example, AI-generated patient narrative videos reduced anxiety, strengthened emotional connectedness, and enriched contextual understanding in nursing simulations (Reed & Dodson, 2024).

Reported challenges of using GenAI in instructional materials development

Accuracy, validity, and reliability of outputs

Several studies highlighted issues of accuracy, validity, and reliability in GenAI-generated outputs. A recurring concern was hallucination or adding information beyond the source text (Al Faraby et al., 2024; Jauhiainen & Guerra, 2023, 2024; Xin, 2024). Al Faraby et al. (2024) also reported declining accuracy across question sequences and imprecision in type-specific questions. Einarsson et al. (2024) found that 23% of reframed problems contained theoretical flaws, with some outputs being overly verbose and requiring editing. Similarly, Lohr et al. (2025) observed that the quality of quizzes often fell below educational standards, with high error rates in answer options and feedback, as well as poorly integrated relational annotations. In clinical assessment, Hudon et al. (2024) noted that scenarios were overly simple or trivial, reducing the reasoning challenge.

In the ML curriculum, Noveski et al. (2024) reported verbose outputs, inconsistent error detection, and unreliable code optimization. Davis (2025) pointed out model limitations in recognizing specific pedagogical terms, resulting in underspecified outputs that required prompt revision and instructional adaptation. Furthermore, GenAI demonstrated limited contextual understanding (particularly when abstract ideas were not accompanied by concrete examples) and struggled to capture complex academic discussions (Garcia-Varela et al., 2025). Contextual limitations also appeared in Arabic language teaching, where GenAI struggled with dialects, grammar, and rhetorical eloquence (*balagha*) (Alkaabi & Almaamari, 2025).

Human readiness

Many GenAI outputs are not ready-to-use and depend heavily on human intervention for filtering and/or refinement. For instance, Einarsson et al. (2024) emphasized the need for expert review to maintain the accuracy of reframed problem statements. Lohr et al. (2025) also reported that GenAI-assisted questions were not immediately ready for use and required prior human filtering. At a larger scale, Bolender et al. (2024) highlighted a strong reliance on Subject Matter Experts to edit and regenerate assessment items. This indicates that educators intending to use GenAI must possess adequate competencies.



Several studies underscored the need to enhance educators' capacity to use GenAI effectively, including in Arabic language teaching (Alkaabi & Almaamari, 2025) and vocational secondary education (Mulyani et al., 2025). In nursing education, Reed & Dodson (2024) stressed the importance of prompting skills and iterative refinement in producing visual materials, suggesting a demand for targeted training. Beyond technical competencies, other studies also highlighted the need to strengthen critical attitudes toward GenAI outputs (Jin et al., 2025; Xin, 2024). Attitudinal readiness is crucial, as Kwack & Im (2023) found that several characteristics of ChatGPT correlated negatively with user attitudes (immediacy, continuity, adjustability, suitability), while expertise did not predict attitudes. These findings point to the risk of superficial or volatile knowledge if the use of GenAI is not properly contextualized.

Technical and operational challenges

Technical and operational challenges include managing complex queries and the need for rigorous testing and validation (Bernal, 2024), integrating AI and developing systems that adapt to teachers' and students' needs (Naatonis et al., 2024), and the requirement of high-quality exemplars for pipelines to generate usable initial drafts (Bolender et al., 2024). On the infrastructure side, glitches and access limitations are reported (Lo, 2025), while constraints on stylistic variation also arise due to API costs (Pesovski et al., 2024).

Prompting burdens and repeated iterations are evident in the production of visual materials, where multiple trials are often needed due to inaccurate or unrealistic images (Reed & Dodson, 2024). In terms of tool capabilities, limitations include the inability to interpret multimodal elements (Xin, 2024) and the inability to generate graphs, which shifted instructional design from graphing activities to the use of quadratic formulas (Davis, 2025).

Ethical issues

In this section, ethical issues are considered broadly to encompass concerns related to bias, stereotypes, data privacy and transparency, and intellectual property. Several studies warned of cultural bias and stereotypes in outputs. Zheng & Stewart (2024) identified WEIRD bias (Western, Educated, Industrial, Rich, and Democratic) and misalignment with cultural principles (Confucian principles) in AI-generated instructional materials for EFL learning. In another English language learning study, P.-A. Nguyen (2024) noted a bias toward Standard English, while Lo (2025) highlighted inaccurate representations of New Englishes.

Stereotype issues also appeared in several studies, including in SCT scenarios in medical education (Hudon et al., 2024), in images generated to illustrate patient backstories in nursing education (Reed & Dodson, 2024), and in visual content produced for language teacher education (Yang & Markauskaite, 2025). Beyond bias and stereotypes, concerns were raised regarding data privacy/security and transparency (Madunić & Sovulj, 2024), as well as copyright and intellectual property issues (Mulyani et al., 2025).

Learning processes and outcomes

Pesovski et al. (2024) found that in the long term, students tended to revert to traditional teaching styles, even though the stylistic variations offered by AI were initially engaging. Lo (2025) emphasized that the "human touch" cannot be fully replaced. In an experiment comparing AI-generated instructional videos with traditionally recorded ones, Xu et al. (2025) reported weaker social presence and lower trust in the AI group, accompanied by

slightly lower motivation and satisfaction. Similarly, Jauhiainen & Guerra (2023, 2024) reported varied perceptions of task difficulty when using AI-generated instructional materials (8.2% rated them as very easy; 5.5% rated them as very difficult), and 13.6% of students indicated that they learned little or nothing.

Discussion

This systematic literature review seeks to synthesize the state of knowledge (Page et al., 2021) in the field of educational technology, with a particular focus on GenAI for instructional materials development. The review of 31 studies that met the inclusion criteria revealed that most of the studies on this topic was conducted in higher education (71%), with the remaining in K-12 contexts (29%). Therefore, future research agendas should increasingly target the K-12 level so that the potential of GenAI in instructional materials development can be leveraged across all levels of education. However, this composition may also have been influenced by the search process, the reliance on Scopus as the sole database, and the inclusion-exclusion criteria applied in this study, which should be considered when interpreting these findings.

The thematic categorization revealed that the types of AI-generated instructional materials included textual and written materials (19 studies), assessment and evaluation materials (12 studies), multimedia materials (11 studies), and game-based and gamified materials (2 studies). Notably, most assessment tasks were also text-based, reinforcing the dominance of textual formats. This trend is consistent with the types of GenAI tools employed, which were largely dominated by large language models (LLMs) such as ChatGPT, LLaMA, Claude, Copilot, and Gemini. Consequently, more research is needed on the use of GenAI for other types of instructional materials, such as videos (Kazanidis & Pellas, 2024) and game-based materials (Ho & Lee, 2023).

The benefits of GenAI in instructional materials development within this corpus were categorized into five themes: (1) efficiency in time, workload, and cost; (2) support for educators; (3) personalization and adaptability; (4) quality and scope of instructional materials; and (5) learning processes and outcomes.

A notable example of efficiency can be seen in the large-scale project aligning 90,000 test items, which demonstrated that GenAI could reduce time requirements by 88% and costs by up to 85%, while maintaining an accuracy rate of 96% and strong expert validation (86–93%) (Bolender et al., 2024). Building on this efficiency potential, GenAI also supports educators by facilitating the search of instructional materials (Kwack & Im, 2023; Romaniuk & Łukasiewicz-Wieleba, 2024) and providing pedagogical suggestions (Xin, 2024), thereby helping to reduce teachers' workload. This aligns with the SWOT analysis conducted by Farrokhnia et al. (2023), which concluded that GenAI tools such as ChatGPT can alleviate teaching workload.

Another benefit of GenAI in instructional materials development is its capacity to facilitate adaptive and personalized learning (Al Faraby et al., 2024; Alkaabi & Almaamari, 2025; Jauhiainen & Guerra, 2023, 2024; Naatonis et al., 2024; Zheng & Stewart, 2024). Integrating GenAI into this approach can reduce educators' workload, increase student engagement, provide scalability for large-scale educational platforms, and enable continuous learning and real-time adjustments (Guettala, Bouekkache, Kazar, & Harous, 2024). Alongside personalization, GenAI also offers the ability to generate diverse content, including high-quality multimedia such as HD videos, synthetic voices, and avatars (Carbonell-Alcocer et al.,

2025), which is expected to further enhance learning processes and outcomes (Byrne, 2024; Reed & Dodson, 2024; Xu et al., 2025). Nevertheless, studies on how GenAI-assisted instructional materials affect learning processes and outcomes remain limited in this corpus, indicating the need for more classroom experiments to observe and measure the wide-ranging impacts of such materials on students' learning processes and outcomes.

Beyond these benefits, a number of challenges in using GenAI for instructional materials development were also identified and grouped into five themes: (1) accuracy, validity, and reliability of outputs; (2) human readiness; (3) technical and operational challenges; (4) ethical issues; and (5) learning processes and outcomes.

One of the primary concerns regarding the accuracy, validity, and reliability of GenAI outputs is the issue of hallucination (Al Faraby et al., 2024; Jauhiainen & Guerra, 2023, 2024; Xin, 2024). AI hallucination, in which the model fabricates incorrect information while sounding highly convincing, remains insufficiently recognized by both educators and learners (Walter, 2024). Hence, AI literacy is essential for all potential users, particularly educators, to ensure that GenAI outputs are accurate, valid, and reliable, and to address the technical and operational challenges that may arise.

With respect to ethical issues, one prominent concern is bias in AI outputs, such as WEIRD bias (Western, Educated, Industrial, Rich, and Democratic), which appeared in moral dilemmatic stories generated by ChatGPT (Zheng & Stewart, 2024). This bias is caused by the training data used in LLMs, which are heavily influenced by WEIRD culture (Atari, Xue, Park, Blasi, & Henrich, 2023; Godwin-Jones, 2024). This finding is also consistent with other systematic literature reviews on GenAI in education, which also reported biases in AI-generated outputs (Ali et al., 2024; Amofa et al., 2025).

Challenges are also evident in relation to learning processes and outcomes. Several studies emphasized the continued importance of the *human touch* in teaching and learning, even though GenAI is capable of producing human-like content (Lo, 2025; Pesovski et al., 2024). This finding supports the statement of Adel, Ahsan, & Davison (2024), who argued that GenAI should complement rather than replace traditional pedagogical methods that rely on human interaction.

The findings of this SLR should be interpreted with caution due to several limitations, including the relatively small corpus size (31 studies), the dominance of higher education contexts (71%), the search strategy, the reliance on Scopus as the sole database, and the inclusion and exclusion criteria applied. As a result, the review may not fully represent the current state of knowledge on the use of GenAI for instructional materials development.

For educators, the findings of this study are expected to encourage the use of various types of GenAI to enhance quality and to develop diverse instructional materials efficiently. GenAI use is expected to foster personalized and contextualized learning, stimulate critical thinking, and create more enjoyable learning experiences, thereby improving student learning outcomes. Nevertheless, educators must also recognize that they need adequate skills to use GenAI effectively, including the ability to address challenges related to output quality, technical limitations, ethical concerns, and the fact that its impact on learning processes and outcomes may not always be positive. Accordingly, institutions bear responsibility for providing regulations, guidelines for GenAI use, technical support, and training programs to help educators design effective prompts, evaluate output quality, and manage one of the most

critical weaknesses of AI, namely hallucinations. In particular, teacher education institutions are expected to equip future teachers with sufficient AI literacy so they can maximize positive benefits while minimizing negative impacts of this technology.

Future research should place greater emphasis on empirical studies, particularly in K-12 contexts, to explore the development of diverse instructional materials, especially non-textual formats such as videos, multimedia, and game-based/gamified resources. It is also important to broaden investigations into the effects of AI-generated instructional materials on learning processes and outcomes, taking into account variations across subject areas and educational levels. Moreover, further studies are needed to identify strategies that can help educators address the challenges of using GenAI in the design, production, and evaluation of instructional materials.

Conclusion

This paper presents a synthesis of research on the use of GenAI for instructional materials development, encompassing the types of materials produced, the tools employed, the reported benefits, and the associated challenges. The findings highlight the dominance of textual and written materials and the extensive use of LLMs/text generation. GenAI demonstrates strong potential to accelerate the development and diversify the formats of instructional materials, with key benefits including efficiency in time, workload, and cost; support for educators; personalization and adaptability; enhanced quality and scope of materials; and positive impacts on learning processes and outcomes. At the same time, the findings underscore critical challenges such as output quality, human readiness, technical–operational barriers, ethical issues, and learning-related impacts that are not always positive. Despite several limitations, this study is expected to encourage educators to strengthen their literacy and skills, institutions to implement appropriate policies and support measures, and researchers to conduct broader and deeper investigations into the use of GenAI in education.

Declarations

Acknowledgments: *The authors would like to express their sincere gratitude to the Directorate of Research and Community Service, Directorate General of Research and Development, Ministry of Higher Education, Science, and Technology of the Republic of Indonesia, for supporting this research. We also thank the anonymous reviewers for their valuable and constructive feedback, which significantly improved the quality of this manuscript.*

Funding: *This research was funded by the Directorate of Research and Community Service, Directorate General of Research and Development, Ministry of Higher Education, Science, and Technology of the Republic of Indonesia, Fiscal Year 2025, under Decree Number 0419/C3/DT.05.00/2025.*

Ethics Statements: *This study is a Systematic Literature Review (SLR) based exclusively on previously published scholarly works. It did not involve human participants, human data, or animal subjects. Therefore, ethical approval was not required.*

Conflict of Interest: *The authors declare that they have no conflict of interest regarding the publication of this paper.*

Informed Consent: *This study did not involve human participants. Therefore, informed consent was not required.*



Data availability: All data analyzed in this study were derived from publicly available published articles. The dataset analyzed during the current study is available from the corresponding author upon reasonable request.

References

- Adel, A., Ahsan, A., & Davison, C. (2024). ChatGPT promises and challenges in education: Computational and ethical perspectives. *Education Sciences*, 14(8), 814. <https://doi.org/10.3390/educsci14080814>
- Agbo, F. J., Olivia, C., Oguibe, G., Sanusi, I. T., & Sani, G. (2025). Computing education using generative artificial intelligence tools: A systematic literature review. *Computers and Education Open*, 9, 100266. <https://doi.org/10.1016/j.caeo.2025.100266>
- Al-Kadi, A., & Ali, J. K. M. (2024). A holistic approach to ChatGPT, Gemini, and Copilot in English learning and teaching. *Language Teaching Research Quarterly*, 43, 155–166. <https://doi.org/10.32038/ltrq.2024.43.09>
- Al Faraby, S., Romadhony, A., & Adiwijaya. (2024). Analysis of LLMs for educational question classification and generation. *Computers and Education: Artificial Intelligence*, 7. <https://doi.org/10.1016/j.caeai.2024.100298>
- Ali, D., Fatemi, Y., Boskabadi, E., Nikfar, M., Ugwuoke, J., & Ali, H. (2024). ChatGPT in teaching and learning: A systematic review. *Education Sciences*, 14(6). <https://doi.org/10.3390/educsci14060643>
- Alkaabi, M. H., & Almaamari, A. S. (2025). Generative AI implementation and assessment in Arabic language teaching. *International Journal of Online Pedagogy and Course Design*, 15(1). <https://doi.org/10.4018/IJOPCD.368037>
- Allehyani, S. H., & Algamdi, M. A. (2023). Digital competences: Early childhood teachers' beliefs and perceptions of ChatGPT application in teaching English as a Second Language (ESL). *International Journal of Learning, Teaching and Educational Research*, 22(11), 343–363. <https://doi.org/10.26803/ijlter.22.11.18>
- Amofa, B., Kamudyariwa, X. B., Fernandes, F. A. P., Osobajo, O. A., Jeremiah, F., & Oke, A. (2025). Navigating the complexity of generative artificial intelligence in higher education: A systematic literature review. *Education Sciences*, 15(7), 826. <https://doi.org/10.3390/educsci15070826>
- Aravantinos, S., Lavidas, K., Voulgari, I., Papadakis, S., Karalis, T., & Komis, V. (2024). Educational approaches with AI in primary school settings: A systematic review of the literature available in Scopus. *Education Sciences*, 14(7), 744. <https://doi.org/10.3390/educsci14070744>
- Atari, M., Xue, M. J., Park, P. S., Blasi, D., & Henrich, J. (2023). *Which humans?* Retrieved from <https://psyarxiv.com/5b26t>.
- Barrot, J. S. (2023). Using ChatGPT for second language writing: Pitfalls and potentials. *Assessing Writing*, 57, 100745. <https://doi.org/10.1016/j.asw.2023.100745>
- Bernal, M. E. (2024). Revolutionizing eLearning assessments: The role of GPT in crafting dynamic content and feedback. *Journal of Artificial Intelligence and Technology*, 4(3), 188–199. <https://doi.org/10.37965/jait.2024.0513>
- Bolender, B., ViSpoel, S., Converse, G., Koprowicz, N., Song, D., & Osaro, S. (2024). Generative AI in K12: Analytics from early adoption. *Journal of Measurement and Evaluation in Education and Psychology*, 15(Special Issue), 361–377. <https://doi.org/10.21031/epod.1539710>
- Byrne, J. (2024). L2 artificial immersion: Piloting the layered language-learning method. *Ubiquitous Learning*, 18(1), 183–199. [Participatory Educational Research \(PER\)](https://doi.org/10.18848/1835-</p>
</div>
<div data-bbox=)

9795/CGP/v18i01/183-199

- Carbonell-Alcocer, A., Sanchez-Acedo, A., Benitez-Aranda, N., & Gertrudix, M. (2025). Impact of generative artificial intelligence on the efficiency, quality, and innovation in the production of Open Educational Resources for MOOCs. *Comunicacion y Sociedad*, e8784. <https://doi.org/10.32870/cys.v2025.8784>
- Cooper, G. (2023). Examining science education in ChatGPT: An exploratory study of generative Artificial Intelligence. *Journal of Science Education and Technology*, 32(3), 444–452. <https://doi.org/10.1007/s10956-023-10039-y>
- Davis, J. D. (2025). Evolving techniques and emerging schemes: Prospective teachers' transformation of ChatGPT. *International Journal of Science and Mathematics Education*, 23, 2181–2198. <https://doi.org/10.1007/s10763-024-10533-8>
- Deroncele-Acosta, A., Sayán-Rivera, R. M., Mendoza-López, A. D., & Norabuena-Figueroa, E. D. (2025). Generative artificial intelligence and transversal competencies in higher education: A systematic review. *Applied System Innovation*, 8(3), 83. <https://doi.org/10.3390/asi8030083>
- Edelson, D. C., Reiser, B. J., McNeill, K. L., Mohan, A., Novak, M., Mohan, L., ... Suárez, E. (2021). Developing research-based instructional materials to support large-scale transformation of science teaching and learning: The approach of the OpenSciEd Middle School Program. *Journal of Science Teacher Education*, 32(7), 780–804. <https://doi.org/10.1080/1046560X.2021.1877457>
- Einarsson, H., Lund, S. H., & Jónsdóttir, A. H. (2024). Application of ChatGPT for automated problem reframing across academic domains. *Computers and Education: Artificial Intelligence*, 6, 100194. <https://doi.org/10.1016/j.caeai.2023.100194>
- Farrokhnia, M., Banihashem, S. K., Noroozi, O., & Wals, A. (2023). A SWOT analysis of ChatGPT: Implications for educational practice and research. *Innovations in Education and Teaching International*, 61(3), 460–474. <https://doi.org/10.1080/14703297.2023.2195846>
- Garcia-Varela, F., Bekerman, Z., Nussbaum, M., Mendoza, M., & Montero, J. (2025). Reducing interpretative ambiguity in an educational environment with ChatGPT. *Computers and Education*, 225, 105182. <https://doi.org/10.1016/j.compedu.2024.105182>
- Godwin-Jones, R. (2024). Distributed agency in language learning and teaching through Generative AI. *Language Learning & Technology*, 28(2), 5–30. <https://doi.org/10.64152/10125/73570>
- Guettala, M., Bourekkache, S., Kazar, O., & Harous, S. (2024). Generative artificial intelligence in education: Advancing adaptive and personalized learning. *Acta Informatica Pragensia*, 13(3), 460–489. <https://doi.org/10.18267/j.aip.235>
- Halaweh, M. (2023). ChatGPT in education: Strategies for responsible implementation. *Contemporary Educational Technology*, 15(2), ep421. <https://doi.org/10.30935/cedtech/13036>
- Ho, W., & Lee, D. (2023). Enhancing engineering education in the Roblox metaverse: Utilizing ChatGPT for game development for electrical machine course. *International Journal on Advanced Science, Engineering and Information Technology*, 13(3), 1052–1058. <https://doi.org/10.18517/ijaseit.13.3.18458>
- Hudon, A., Kieppura, B., Pelletier, M., & Phan, V. (2024). Using ChatGPT in psychiatry to design script concordance tests in undergraduate medical education: Mixed methods study. *JMIR Medical Education*, 10, e54067. <https://doi.org/10.2196/54067>
- 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>
- Jauhiainen, J. S., & Guerra, A. G. (2024). Generative AI and education: Dynamic



- personalization of pupils' school learning material with ChatGPT. *Frontiers in Education*, 9, 1288723. <https://doi.org/10.3389/feduc.2024.1288723>
- Jin, F., Peng, X., Sun, L., Song, Z., Zhou, K., & Lin, C.-H. (2025). Knowledge (co-)construction among artificial intelligence, novice teachers, and experienced teachers in an online professional learning community. *Journal of Computer Assisted Learning*, 41(2), e70004. <https://doi.org/10.1111/jcal.70004>
- Kazanidis, I., & Pellas, N. (2024). Harnessing generative artificial intelligence for digital literacy innovation: A comparative study between early childhood education and computer science undergraduates. *AI (Switzerland)*, 5(3), 1427–1445. <https://doi.org/10.3390/ai5030068>
- Kwack, H. R., & Im, E. A. (2023). A study of pre-service teachers' perceptions of the use of ChatGPT in practical gardening training. *Journal of People, Plants, and Environment*, 26(6), 693–709. <https://doi.org/10.11628/ksppe.2023.26.6.693>
- Labadze, L., Grigolia, M., & Machaidze, L. (2023). Role of AI chatbots in education: Systematic literature review. *International Journal of Educational Technology in Higher Education*, 20(1), 56. <https://doi.org/10.1186/s41239-023-00426-1>
- Leite, B. S. (2024). Generative Artificial Intelligence in chemistry teaching: ChatGPT, Gemini, and Copilot's content responses. *Journal of Applied Learning and Teaching*, 7(2), 190–204. <https://doi.org/10.37074/jalt.2024.7.2.13>
- Liu, J., Sihes, A. J. B., & Lu, Y. (2025). How do generative artificial intelligence (AI) tools and large language models (LLMs) influence language learners' critical thinking in EFL education? A systematic review. *Smart Learning Environments*, 12(1), 48. <https://doi.org/10.1186/s40561-025-00406-0>
- Lo, A. W. T. (2025). The educational affordances and challenges of generative AI in Global Englishes-oriented materials development and implementation: A critical ecological perspective. *System*, 130, 103610. <https://doi.org/10.1016/j.system.2025.103610>
- Lohr, D., Berges, M., Chugh, A., Kohlhase, M., & Müller, D. (2025). Leveraging large language models to generate course-specific semantically annotated learning objects. *Journal of Computer Assisted Learning*, 41(1), e13101. <https://doi.org/10.1111/jcal.13101>
- Madunić, J., & Sovulj, M. (2024). Application of ChatGPT in information literacy instructional design. *Publications*, 12(2), 11. <https://doi.org/10.3390/publications12020011>
- Marzano, D. (2025). Generative artificial intelligence (GAI) in teaching and learning processes at the K-12 level: A systematic review. *Technology, Knowledge and Learning*. <https://doi.org/10.1007/s10758-025-09853-7>
- Mulyani, H., Istiaq, M. A., Shauki, E. R., Kurniati, F., & Arlinda, H. (2025). Transforming education: exploring the influence of Generative AI on teaching performance. *Cogent Education*, 12(1). <https://doi.org/10.1080/2331186X.2024.2448066>
- Naatonis, R. N., Rusijono, Jannah, M., & Malahina, E. A. U. (2024). Evaluation of problem based gamification learning (PBGL) model on critical thinking ability with artificial intelligence approach integrated with ChatGPT API: An experimental study. *Qubahan Academic Journal*, 4(3), 485–520. <https://doi.org/10.48161/qaj.v4n3a919>
- Nguyen, P.-A. (2024). Evaluating AI-generated language as models for strategic competence in English language teaching. *IAFOR Journal of Education*, 12(3), 325–349. <https://doi.org/10.22492/IJE.12.3.13>
- Nikolic, S., Sandison, C., Haque, R., Daniel, S., Grundy, S., Belkina, M., ... Neal, P. (2024). ChatGPT, Copilot, Gemini, SciSpace and Wolfram versus higher education assessments: an updated multi-institutional study of the academic integrity impacts of Generative Artificial Intelligence (GenAI) on assessment, teaching and learning in engineering.

- Australasian Journal of Engineering Education*, 29(2), 126–153. <https://doi.org/10.1080/22054952.2024.2372154>
- Noveski, G., Jeroncic, M., Velard, T., Kocuvan, P., & Gams, M. (2024). Comparison of large language models in generating machine learning curricula in high schools. *Electronics (Switzerland)*, 13(20), 4109. <https://doi.org/10.3390/electronics13204109>
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ... Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*, 372, n71. <https://doi.org/10.1136/bmj.n71>
- Pesovski, I., Santos, R., Henriques, R., & Trajkovik, V. (2024). Generative AI for customizable learning experiences. *Sustainability*, 16(7). <https://doi.org/10.3390/su16073034>
- Powell, W., & Courchesne, S. (2024). Opportunities and risks involved in using ChatGPT to create first grade science lesson plans. *PLOS ONE*, 19(6), e0305337. <https://doi.org/10.1371/journal.pone.0305337>
- Prins, G. T., Bulte, A. M. W., & Pilot, A. (2018). Designing context-based teaching materials by transforming authentic scientific modelling practices in chemistry. *International Journal of Science Education*, 40(10), 1108–1135. <https://doi.org/10.1080/09500693.2018.1470347>
- Rahman, M. M., & Watanobe, Y. (2023). ChatGPT for education and research: Opportunities, threats, and strategies. *Applied Sciences*, 13(9), 5783. <https://doi.org/10.3390/app13095783>
- Ranuharja, F., Ganefri, Rizal, F., Langeveldt, D., Ejjami, R., Torres-Toukoumidis, A., & Jalinus, N. (2025). Relevance and impact of generative AI in vocational instructional material design: A systematic literature review. *Salud, Ciencia y Tecnologia*, 5. <https://doi.org/10.56294/saludcyt20251336>
- Reed, J. M., & Dodson, T. M. (2024). Generative AI backstories for simulation preparation. *Nurse Educator*, 49(4), 184–188. <https://doi.org/10.1097/NNE.0000000000001590>
- Romaniuk, M. W., & Łukasiewicz-Wieleba, J. (2024). Generative artificial intelligence in the teaching activities of academic teachers and students. *International Journal of Electronics and Telecommunications*, 70(4), 1043–1048. <https://doi.org/10.24425/ijet.2024.152092>
- Shree, D., Singh, R. K., Paul, J., Hao, A., & Xu, S. (2021). Digital platforms for business-to-business markets: A systematic review and future research agenda. *Journal of Business Research*, 137, 354–365. <https://doi.org/10.1016/j.jbusres.2021.08.031>
- Walter, Y. (2024). Embracing the future of Artificial Intelligence in the classroom: the relevance of AI literacy, prompt engineering, and critical thinking in modern education. *International Journal of Educational Technology in Higher Education*, 21(1), 15. <https://doi.org/10.1186/s41239-024-00448-3>
- Wang, M., Tlili, A., Khribi, M. K., Lo, C. K., & Huang, R. (2025). Generative artificial intelligence in special education: A systematic review through the lens of the mediated-action model. *Information Development*, 0(0). <https://doi.org/10.1177/02666669251335655>
- Wang, P., Jing, Y., & Shen, S. (2025). A systematic literature review on the application of generative artificial intelligence (GAI) in teaching within higher education: Instructional contexts, process, and strategies. *The Internet and Higher Education*, 65, 100996. <https://doi.org/10.1016/j.iheduc.2025.100996>
- Wardat, Y., Tashtoush, M. A., AlAli, R., & Jarrah, A. M. (2023). ChatGPT: A revolutionary tool for teaching and learning mathematics. *Eurasia Journal of Mathematics, Science and Technology Education*, 19(7), em2286. <https://doi.org/10.29333/ejmste/13272>
- Xin, J. J. (2024). Investigating EFL teachers' use of Generative AI to develop reading



- materials: A practice and perception study. *Language Teaching Research*.
<https://doi.org/10.1177/13621688241303321>
- Xu, T., Liu, Y., Jin, Y., Qu, Y., Bai, J., Zhang, W., & Zhou, Y. (2025). From recorded to AI-generated instructional videos: A comparison of learning performance and experience. *British Journal of Educational Technology*, 56, 1463–1487. <https://doi.org/10.1111/bjet.13530>
- Yang, H., & Markauskaite, L. (2025). Fostering language student teachers' transformative agency for embracing GenAI: A formative intervention. *Teaching and Teacher Education*, 159, 104980. <https://doi.org/10.1016/j.tate.2025.104980>
- Zhang, X. (2025). Incorporating AI tools into database classes. *Journal of Information Systems Education*, 36(1), 37–52. <https://doi.org/10.62273/GKZI2477>
- Zheng, Y. D., & Stewart, N. (2024). Improving EFL students' cultural awareness: Reframing moral dilemmatic stories with ChatGPT. *Computers and Education: Artificial Intelligence*, 6, 100223. <https://doi.org/10.1016/j.caeai.2024.100223>