AIER
ARTIFICIAL INTELLIGENCE
in EDUCATIONAL RESEARCH
E-ISSN: 3023-8048

I Tutor: A Systematic Review of Artificially Intelligent Tutors for the Classroom[1]

Allyson Hauptman[2]

Review Article **Received:** 8 Oct. 2025 **Accepted:** 29 Nov. 2025

Abstract: This systematic review aims to understand what research has been done on the use of AI as a classroom tutor and how that body of work should shape future research. A systematic review was conducted using key term searches in four major, peer-reviewed journal databases for relevant research and categorized them by research type, AI type, assistance type, and education level and analyzed them for study measures and key findings. After applying inclusion and exclusion criteria, a total of 34 research articles were analyzed for key findings. The review found that most studies have focused on chatbots supporting undergraduate level education. Key findings show that AI tutors may be most useful for students beginning at a lower level of expertise than their peers, and the use of techniques such as RAG may significantly enhance AI's usefulness as a tutor with niche and higher-level subjects. This systematic review uniquely considers the last decade of research on AI use as a classroom tutor and provides insights into where future research on the subject may be most useful and impactful to the field of education.

Keywords: Generative AI, GenAI, intelligent tutoring.

[1] doi: https://doi.org/10.5281/zenodo.17763699
[2] Author, United States Military Academy, USA, ORCID: 0000-0001-7785-5996

Introduction

In 2018, the world was introduced to Generative AI (GenAI) by the OpenAI project. In 2022, it would become an easily accessible tool with the public release of ChatGPT (Wu et al., 2003). Since then, educational institutions have struggled to determine how to best leverage artificial intelligence (AI) in the classroom. Initially, GenAI was mostly used to quickly summarize and fact-check information. More recently, students have discovered its use as a ghost writer and collaborator. Researchers have begun experimenting with the possibility of using AI chatbots as a tutor, where the chatbot can answer student questions based on the course syllabus and material (Chen et al., 2023;). There is a wide variety of uncertainties regarding the extent to which AI, both GenAI and other forms of AI, can be used as a classroom tutor. These range from the AI's understanding of course material to its ability to effectively communicate with and guide a student.

A few previous literature reviews have sought to understand how AI can be used to enhance classroom learning for K-12 education, as either an in-classroom tools (Lee & Kwon, 2024) or as intelligent tutoring systems (Létourneau et al., 2025), while other studies have focused on challenges and opportunities for incorporating AI into the classroom (Dimitriadou & Laniti, 2023; Jaramillo & Chiappe, 2024). What is currently lacking, and what this systematic review aimed to understand, is an evaluation of what empirical research exists on the benefits and limitations of AI as a classroom tutor. To do this, the review conducted a key term search for relevant research articles in four of the most extensive peer-reviewed journal databases and analyzed them for important research methods, measures, and findings. The results of this review provide insight into how AI can effectively be used as a tutor in classroom environments and the existing research gaps that should be explored to capitalize on those opportunities.

Background

The GenAI Boom

Recent advancements in the complexity of Large Language Models (LLMs) have enabled the development of a type of AI referred to as generative AI, or GenAI. The definition of just what an LLM must produce to be considered GenAI is not universal, but it is widely accepted that the model must generate "high-quality, human-like material" (Penalvo et al., 2023). The GenAI that is largely behind the GenAI boom, ChatGPT, is actually OpenAI's GPT-3 model. GPT and GPT-2 were released in 2018 (Heaven, 2023). So what made GPT-3 different? For the public, it was its ground-breaking ability to answer questions and generate summaries like a human being.

Within years, GenAI has evolved from being an exciting commodity to a necessary productivity advantage. GenAI models can now not only answer questions but also analyze trends, create images, and write code (Bharti et al., 2025). The speed at which GenAI has ingrained itself in everyday life has produced a new research focus on how this disruptive technology should (or should not) be utilized in a variety of sectors, such as governance, healthcare, and education (Delios et al., 2024).

The Promises of Intelligent Tutoring

Intelligent tutoring systems (ITSs) evolved from other computer-assisted instruction methods as a way in which AI could support education (Nwana, 1990). Traditionally, ITSs most often supported computer science fields and provided students with near real-time feedback on their products and processes. One of the most promising AI-enabled aspects of these systems is the ability of the system to learn and provide adaptive instruction and feedback to the users (Mousavinasab et al., 2021).

GenAI is revolutionizing this concept and driving research into how LLMs such as GPT-4 can provide students with personalized learning experiences. The ways in which these systems can produce personalized instruction include question generation for quizzes, immediate evaluation and feedback, and customized learning content (Maity and Deroy, 2024). Such systems are exciting for educators tasked with teaching students with various aptitudes, backgrounds, and learning styles. A study by McKinsey & Company determined that one of the sectors most poised for disruption by AI-powered tools such as these is education (Bughin et al., 2018). Yet, despite these possibilities, no research has yet tied together overall trends and results of the use of AI tutors in the classroom. Thus, the research questions that this review aims to answer are:

RQ1: *What research has been done on the use of AI tutor's in the classroom?*

RQ2: *What are the implications for its use in the near future?*

Methodology

The first step in completing this review was to select the number and appropriate databases within which to conduct a key term search. Four databases were selected:

(1) Institute of Electrical and Electronics Engineers (IEEE) Xplore
(2) Association for Computing Machinery (ACM) Digital Library
(3) Science Direct
(4) Springer Nature Journals

These databases were chosen due to their focus on applied technology and education research. These databases also represent four of the most extensive peer-reviewed article databases available, thus ensuring a thorough, comprehensive review of the literature. Within these databases, in April of 2025, the review initially searched the following key terms one by one using the database Advanced Search tools. Terms were searched one at a time by the review author in order to associate the results with the proper term:

(1) Generative AI tutor

(2) GenAI tutor

(3) Artificially intelligent tutor

However, due to the low number of returned articles, the search was expanded to include a fourth key term:

(4) AI tutor

The addition of this term did not change the scope of the review, as it was a variation on one of the initially searched terms, but it did mean that the application of exclusion criteria would be even more important to prevent scope creep.

In conducting the searches within these four databases, the queries were limited to peer-reviewed research articles published in or after 2018. The reason for this is to focus on the AI boom, following the release of OpenAI's GPT and GPT-2 models (Roumeliotis and Tselikas, 2023). The search terms were also required to be listed within the articles' abstracts, thus ensuring that the focus of the paper itself was on the review subject, as opposed to just being mentioned within the text. Following the initial database searches, all the collected articles were reviewed against two additional inclusion criteria. (1) The subject of the paper is an AI agent or system, and (2) the purpose of the AI is to provide student assistance. The search process is illustrated below in Figure 1.

Analysis

Once the search was complete, the reviewer conducted a close analysis of all the articles to extract a number of key elements. First, the articles were labeled as either theoretical research or empirical research. Second, they were categorized by what type of AI was discussed and/or used in the study: Chatbot, Recommender System, Evaluation System, Robot, or Virtual Reality. Third, they were categorized by what type of assistance was given to the student by the AI: Question & Answer, Content Delivery, Feedback or Real-time Assistance. Finally, they were categorized by the level of education of the assisted students: K-12, Undergraduate, Graduate/Postgraduate, or Vocational.

Next, the reviewer extracted the measures discussed and/or used in the research articles. These included empirical measures, or those performed on a subject as a result of an experimental manipulation, and perceptual measures, or those reported by a subject as a result of an experimental manipulation. Finally, the reviewer conducted a textual analysis of the study results to better understand the implications of the research.

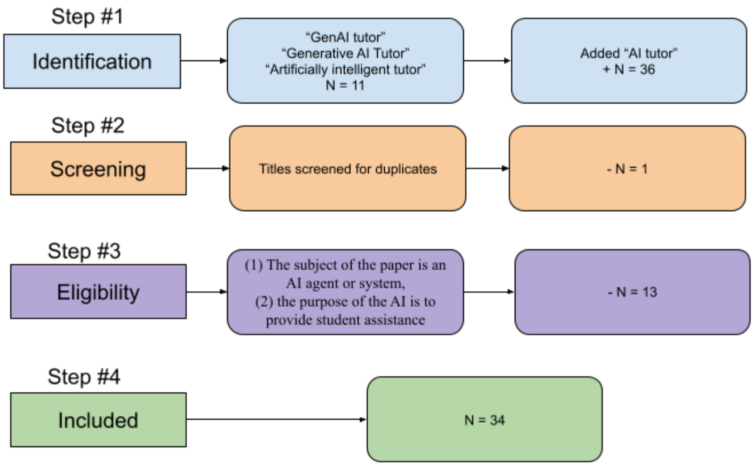

Figure 1: Search Process Diagram

Findings

Search Results

The initial key term search within the four listed databases returned a total of 47 articles for analysis, with most of the articles coming from Springer Nature Journals and IIEE Explore (Table 1). Markedly, with a total of 36 returns, the "AI tutor" search term produced the most articles for analysis and was also the only term to produce any results in ACM Journals.

Table 1.
Key Term Search Results

	IEEE	ACM	Science Direct	Springer Nature	Totals
GenAI tutor	6	0	1	1	8
Generative AI tutor	0	0	0	1	1
Artificially intelligent tutor	0	0	1	1	2
AI tutor	13	4	1	18	36
Totals	19	4	3	21	47

The full initial results shown in Table 1 support the choice to include a fourth search term, as without it the results would have been a mere 11 articles. Once the 47 articles were extracted from the databases, the reviewer began to assess them according to the inclusion criteria. Thirteen articles were removed from the review following the application of inclusion criteria (Table 2), to include one article that was removed as a duplicate. This left 34 articles for consideration by the review. Notably, the application of inclusion criteria resulted in the removal of all articles returned from the key terms "generative AI tutor" and "artificially intelligent tutor," meaning that the actual analysis covers only articles produced by two, not four, key terms.

Table 2.
Results of Applying Exclusion Criteria

	IEEE	ACM	Science Direct	Springer Nature	Totals
GenAI tutor	4	0	1	0	*5*
Generative AI tutor	0	0	0	0	*0*
Artificially intelligent tutor	0	0	0	0	0
AI tutor	11	4	1	13	29
Totals	15	4	2	13	34

Article Categorizations

Most of the articles were categorized as empirical research, with only three articles being purely theoretical (Gan et al., 2019; Kloos et al., 2024; Shahri et al., 2024). Of those 31 empirical research articles, the majority utilized quantitative and mixed methods. Of note is that one article utilized a participatory design study that yielded purely qualitative data (Holstein et al., 2019). The largest educational focus area of the studies was undergraduate education. This may be related to the phenomenon of convenience sampling, as undergraduate students are generally willing and able to serve as a sample for graduate and postgraduate students at research universities (Peterson and Merunka, 2014). Concerningly, the next largest focus population was unspecified, meaning that the research either a) included a mixed population or b) did not include participant data. Overwhelmingly, Chatbots were the most common form of AI within the research, being present in 21 of the articles. This is understandable based on the ongoing generative AI boom. While most of the chatbots used are publicly available models, such as ChatGPT, some were not. Two of the studies developed custom models (Bonde, 2024; Yang et al., 2024), while three studies modified existing models utilizing specific course materials from which to deliver answers and content to the students using techniques such as Retrieval-Augmented Generation (RAG) (Golde and Geng, 2024; Lui et al., 2024; Soliman et al., 2024). The next most common forms of AI were recommender systems and evaluation systems. In the former case, the AI was most often responsible for providing course content to a student based on their needs (Baillifard, 2024; Gan et al., 2019; Kim et al., 2020; Madhuri et al., 2024). In the latter case, the AI was most often responsible for evaluating a student's exam or project and providing immediate feedback (Abduljabbar et al., 2022; Karumbaiah et al., 2023; Khabarov et al., 2024).

Considering that the most common type of AI used in the research was a chatbot, it follows that the most common type of assistance was question and answer between the student and the AI. The next most prominent type of assistance given by the AI was content delivery, where the AI provided lessons or practice exams to the student. This was closely followed in prominence by feedback, where the AI reviewed a student's product and provided feedback on its accuracy. These products included program code, graded assignments, and exams. A one-off type of assistance provided to a student was real-time assistance, where an AI agent guided the student in a virtual reality environment (Vijay et al., 2020).

Measures

A total of 25 empirical measures were extracted from the articles (Table 3). While many of these measures were context-dependent measures unique to one study, there were a few stand-out repeated measures. Accuracy of the AI was the most repeated measure and was used in eight studies. This measure generally related to how accurate an AI agent's response was to a student's question but also includes the accuracy of its evaluation of a student's completed product when reviewed by a human instructor.

Student grade was the next most repeated measure and was used in seven studies. Note, this measure only considers a one-time grade, while a separate measure was extracted for grade improvement. Another performance-related metrics, skill completion was the third most repeated measure. In these studies, students were assessed for completing multiple individual skills, as opposed to one overall assignment or exam. Other repeated measures included, in order of prominence: times the student asked the AI, grade improvement, question type, and student attempts.

A total of 15 perceptual measures were extracted from the articles (Table 4). These measures were derived by the researchers through surveys and interviews with both student and teacher subjects. Two perceptions dominated the analysis. The most prominent measurement was the perceived usefulness of the AI. This perception was measured according to multiple usefulness scales on both student and teacher surveys and reflects its use as a common measurement in human factors research (Adams et al., 1992).

Table 3.
Empirical Measures Used in the Research Articles

Measurement	Frequency	Articles
Accuracy (of the AI)	8	Frank et al (2024) Makharia et al (2024) Kim et al. (2020) Madhuri et al. (2024) Hurt et al. (2020) Bonde (2024) Looi et al. (2025) Soliman et al. (2024)
Grade	7	Huo et al. (2024) Llanos et al. (2024) Hurt et al. (2020) Paudyal et al. (2020) Cheng et al. (2025) Baillifard et al. (2025) Gold et al. (2024)
Student Skill Completion	5	Vijay et al. (2020) Borchers et al. (2024) Sarshartehrani et al. (2024) Banerjee et al. (2020) Mailach et al. (2025)
Times Asked (by the student)	4	Huo et al. (2024) Frank et al. (2024) Borchers et al. (2024) Cheng et al. (2025)
Question Type	3	Huo et al. (2024) Cheng et al. (2025) Mailach et al. (2025)
Grade Improvement	3	Llanos et al. (2024) Borchers et al. (2024) Baillifard et al.(2024)
Student Attempts	2	Hurt et al.(2024) Borchers et al. (2024)
Content Uploaded	1	Huo et al. (2024)
Questions Resolved	1	Huo et al. (2024)
Pascal's Triangle Exercise	1	Frankford et al (2024)
Coherence (of the AI)	1	Makharia et al. (2024)
Degree of Assistance	1	Makharia et al. (2024)
No of Recommendations	1	Kim et al. (2020)
Preference Match	1	Kim et al. (2020)
Usage Rate	1	Llanos et al. (2024)
Tardiness (of the student)	1	Llanos et al. (2024)
Course Retention	1	Llanos et al. (2024)
Error Rate (of the student)	1	Vijay et al. (2020)
Teacher Behavior	1	Borchers et al. (2024)
Time to Complete (by the student)	1	Yang et al. (2021)
Question Depth	1	Cheng et al. (2025)
Question Mechanism	1	Cheng et al. (2025)
Objective Alignment	2	Gold et al. (2025) Yang et al. (2024)
Change in Response Quality	3	Looi et al. (2025) Yang et al. (2024) Kim et al. (2020)
Student engagement	2	Sarshartehrani et al. (2024) Karumbaiah et al. (2023)

The second most common perceptual measure was satisfaction with the AI. This perception refers to the satisfaction of both the student receiving the assistance and the instructor evaluating the assistance given by the AI. Other repeated perceptual measures included, in order of prominence: Understanding of the AI, helpfulness of the AI, ease of use of the AI, and student engagement with the material.

Table 4.
Perceptual Measures Used in the Research Articles

Measure	Frequency	Articles
Usefulness (of the AI)	6	Lui et al. (2024) Yang et al. (2024) Frankford et al. (2024) Belkacem & Hireche (2024) Chen et al. (2023) Karumbaiah et al. (2023)
Satisfaction (with the AI)	5	Huo et al. (2024) Llanos et al. (2024) Yang et al. (2021) Belkacem & Hireche (2024) Looi et al. (2025)
Understanding (of the AI)	3	Frank et al. (2024) Belkacem & Hireche (2024) Looi et al. (2025)
Ease of Use (of the AI)	2	Lui et al. (2024) Frankford et al. (2024)
Enjoyment (with the AI)	2	Lui et al. (2024) Belkacem & Hireche (2024)
Student Engagement (with the material)	2	Tan et al. (2024) Belkacem & Hireche (2024)
Student Engagement (with the material)	2	Tan et al. (2024) Belkacem & Hireche (2024)
Helpfulness (of the AI)	2	Gold et al. (2025) Looi et al. (2025)
Pace	1	Belkacem & Hireche (2024)
Reliance (on the AI)	1	Looi et al. (2025)
Quality (of the AI support)	1	Chen et al. (2023)
Credibility (of the AI)	1	Chen et al. (2023)
Controllability (of the content)	1	Chen et al. (2023)
Supportiveness (of the AI)	1	Holstein et al. (2019)
Needs anticipation (by the AI)	1	Holstein et al. (2019)

Discussion

Trends and Implications

There are a few notable trends highlighted in Section IV that this section will address. First, it is important to acknowledge that only one of the studies conducted an in-depth qualitative analysis of how AI tutors affect human perceptions, and even this study was isolated to teachers' perceptions of how an AI tutor should be designed and its purpose in the classroom (Holstein et al., 2019). This participatory design study is useful in understanding the desires of educators, but it provides no insight into how a specific AI tutor impacts human user perceptions. From a human factors standpoint, this is a huge gap in the literature, as research on human interaction with assistants, specifically chatbots, shows that a user's perception of the AI agent's qualities significantly affects their likelihood to accept its assistance and seek it again in the future (Cheng et al., 2022). There are a decent number of studies that have aimed to capture this through quantitative means, but because human factors research is so deeply rooted in context and individual differences, it is difficult to derive causation from purely quantitative measures (Maxwell, 2012).

There is an intense focus in the literature on the related measures of accuracy, performance, satisfaction, and usefulness. This makes a lot of sense, as the first concern over using a tutor is how well it helps the student's mastery of the intended material. It is interesting that the measure of "grade" was more prominent than "grade improvement," as it stands to reason that the degree to which an AI tutor was successful was how well it increased a student's knowledge. In the former case, you may have students who are starting from a higher or lower aptitude, regardless of the AI's tutelage. In fact, one of the studies observed the phenomenon where the AI tutor was more effective for the "lower learning rate" student test population, indicating the AI tutors may be even more useful for students starting with a lower aptitude for a subject than those with a higher aptitude (Borchers et al., 2024). It was unsurprising that most of the studies focused on Q&A chatbots, where a student engages with the tutor in the same way as they would engage with any of the popular Generative AI models. Where some of these studies seem to have the most promise was in the cases where the models were tailored to first- or only- utilize course materials in answering the student's questions (Gold et al., 2024; Lui et al., 2024; Soliman et al., 2024). The use of RAG can significantly increase the effectiveness of an AI tutor for a student, because it is highly effective for generating answers for less-popular, well known knowledge sets (Soudani et al., 2024). In essence, while the World Wide Web may be a great data set for an AI agent to produce an answer about a well-known film, it is less optimal for producing an answer for a student studying a niche discipline. This also means that such methods may be even more relevant for higher education than for K-12, as widely studied foundational courses are better represented in publicly available models.

One of the final aspects that should receive some attention is the possibilities of AI tutors in virtual and augmented reality. One of the studies that seems particularly promising explored its use in engineering education (Vijay et al., 2020). The concept that AI tutors could assist students in understanding hands-on disciplines from outside of the laboratory environment has extraordinary implications for hybrid and remote learning. If an AI tutor could deliver instruction with real-time feedback to students studying disciplines grounded in physical skills, it would not only enable reinforced learning outside of dedicated class time but also help many disciplines tackle teacher shortages (Lawal, 2024). Additionally, only two of the studies were focused on vocational tutoring, and the need for hands-on instruction may be a reason for this research gap. The possibility of delivering AI instruction through virtual or augmented reality is uniquely promising for the continual education of workers in hands-on environments.

Limitations and Future Research

This review has a few notable limitations. First, to focus on the post-ChatGPT generative AI boom, the review only includes research from 2018 and beyond. The technologies that enabled the 2018 boom existed years before, and so there may be additional relevant studies that preceded the inclusion criteria. Second, as evidenced in two of the search terms ending up irrelevant, it is likely more articles could have been included by isolating and searching on more of the terms (i.e. searching just for "generative"). While such searches would have likely yielded some more articles for analysis, they also would have exploded the number of studies for subjection to exclusion criteria, which is why it was discarded as an option for this review.

As noted in the discussion, this review presents several opportunities for future research on AI tutors. Foremost, there is a distinct lack of qualitative research on how AI tutors influence student perceptions, perceptions that will inevitably affect the acceptance and use of the tutors in the long term. Second, none of the studies included in the analysis were longitudinal and, given how important grade improvement is to the efficacy of a tutor, studies that assess an AI tutor over longer periods are essential to improving their design and implementation. Finally, as virtual and augmented reality tools become more commonly available in classroom environments, researchers should further investigate how AI tutors can be employed in them to provide tutoring for hands-on, context-dependent disciplines.

Conclusion

This systematic review conducted a search in four of the most extensive peer-reviewed journal databases on key terms related to artificially intelligent tutors to understand how they have been studied and utilized in classroom environments in the post-ChatGPT generative AI boom. Following the application of inclusion and exclusion criteria, it produced 34 articles for analysis. This analysis included the classification of types of AI, types of tutoring assistance given, student education level, and the effects and perceptions measured. The results of the analysis revealed a number of interesting trends and opportunities for future AI research, including important measures, the types of AI utilized, and study methods.

Acknowledgement

Ethics statement: In this study, I affirm compliance with the rules outlined in the "Higher Education Institutions Scientific Research and Publication Ethics Directive" and assert that none of the "Actions Against Scientific Research and Publication Ethics" have been undertaken. Furthermore, we declare that there is no conflict of interest among the authors, that all authors have contributed to the study, and that full responsibility for any ethical violations rests with the article authors.

Funding: This research received no funding.

Declaration of competing interest: There is no conflicts of interest related to this submission.

References

Abduljabbar, A., Gupta, N., Healy, L., Kumar, Y., Li, J. J., & Morreale, P. (2022). *A self-served AI tutor for growth mindset teaching*. In *2022 5th International Conference on Information and Computer Technologies (ICICT)* (pp. 55–59). IEEE.

Adams, D. A., Nelson, R. R., & Todd, P. A. (1992). Perceived usefulness, ease of use, and usage of information technology: A replication. *MIS Quarterly*, *16*(2), 227–247.

Baillifard, A., Gabella, M., Lavenex, P. B., & Martarelli, C. S. (2025). Effective learning with a personal AI tutor: A case study. *Education and Information Technologies, 30*(1), 297–312.

Banerjee, A., Lamrani, I., Hossain, S., Paudyal, P., & Gupta, S. K. S. (2020). AI enabled tutor for accessible training. In *Artificial Intelligence in Education: 21st International Conference, AIED 2020, Ifrane, Morocco, July 6–10, 2020, Proceedings, Part I* (Vol. 21, pp. 29–42). Springer International Publishing.

Belkacem, A. N., & Hireche, A. (2024). Exploring student engagement and learning preferences: A comparative study between virtual- and robot-based tutors. *Measurement: Sensors*, *101*, 704.

Bharti, I., Chauhan, K., & Aggarwal, P. (2025). Generative AI: Next frontier for competitive advantage. In *Enhancing Communication and Decision-Making With AI* (pp. 1–36). IGI Global.

Bonde, L. (2024). A generative artificial intelligence-based tutor for personalized learning. In *2024 IEEE SmartBlock4Africa* (pp. 1–10). IEEE.

Borchers, C., Wang, Y., Karumbaiah, S., Ashiq, M., Shaffer, D. W., & Aleven, V. (2024). Revealing networks: Understanding effective teacher practices in AI-supported classrooms using transmodal ordered network analysis. In *Proceedings of the 14th Learning Analytics and Knowledge Conference* (pp. 371–381).

Bughin, J., Seong, J., Manyika, J., Chui, M., & Joshi, R. (2018). *Notes from the AI frontier: Modeling the impact of AI on the world economy*. McKinsey Global Institute.

Chen, Y., Jensen, S., Albert, L. J., Gupta, S., & Lee, T. (2023). Artificial intelligence (AI) student assistants in the classroom: Designing chatbots to support student success. *Information Systems Frontiers, 25*(1), 161–182.

Cheng, X., Zhang, X., Cohen, J., & Mou, J. (2022). Human vs. AI: Understanding the impact of anthropomorphism on consumer response to chatbots from the perspective of trust and relationship norms. *Information Processing & Management, 59*(3), 102940.

Cheng, Y., Fan, Y., Li, X., Chen, G., Gašević, D., & Swiecki, Z. (2025). Asking generative artificial intelligence the right questions improves writing performance. *Computers and Education: Artificial Intelligence, 8*, 100374.

Delios, A., Tung, R. L., & van Witteloostuijn, A. (2024). How to intelligently embrace generative AI: The first guardrails for the use of GenAI in IB research. *Journal of International Business Studies*, 1–10.

Dimitriadou, E., & Lanitis, A. (2023). A critical evaluation, challenges, and future perspectives of using artificial intelligence and emerging technologies in smart classrooms. *Smart Learning Environments, 10*(1), 12.

Frank, L., Herth, F., Stuwe, P., Klaiber, M., Gerschner, F., & Theissler, A. (2024). Leveraging GenAI for an intelligent tutoring system for R: A quantitative evaluation of large language models. In *2024 IEEE Global Engineering Education Conference (EDUCON)* (pp. 1–9). IEEE.

Frankford, E., Sauerwein, C., Bassner, P., Krusche, S., & Breu, R. (2024). AI-tutoring in software engineering education. In *Proceedings of the 46th International Conference on Software Engineering: Software Engineering Education and Training* (pp. 309–319).

Gan, W., Sun, Y., Ye, S., Fan, Y., & Sun, Y. (2019). AI-tutor: Generating tailored remedial questions and answers based on cognitive diagnostic assessment. In *2019 6th International Conference on Behavioral, Economic and Socio-Cultural Computing (BESC)* (pp. 1–6). IEEE.

Gold, K., & Geng, S. (2024). On the helpfulness of a zero-shot Socratic tutor. In *International Conference on Artificial Intelligence in Education Technology* (pp. 154–170). Springer Nature Singapore.

Heaven, W. D. (2023). ChatGPT is everywhere. Here's where it came from. *MIT Technology Review*, *10*.

Holstein, K., McLaren, B. M., & Aleven, V. (2019). Designing for complementarity: Teacher and student needs for orchestration support in AI-enhanced classrooms. In *Artificial Intelligence in Education: 20th International Conference, AIED 2019, Chicago, IL, USA, June 25–29, 2019, Proceedings, Part I* (Vol. 20, pp. 157–171). Springer International Publishing.

Huo, H., Ding, X., Guo, Z., Shen, S., Ye, D., Pham, O., Milne, D., Mathieson, L., & Gardner, A. (2024). Accelerating learning with AI: Improving students' capability to receive and build automated feedback for programming courses. In *2024 World Engineering Education Forum–Global Engineering Deans Council (WEEF-GEDC)* (pp. 1–9). IEEE.

Hurt, J., Runyon, M., Hammond, T. A., & Linsey, J. S. (2020). A study on the impact of a statics sketch-based tutoring system through a truss design problem. In *2020 IEEE Frontiers in Education Conference (FIE)* (pp. 1–7). IEEE.

Jaramillo, J. J., & Chiappe, A. (2024). The AI-driven classroom: A review of 21st century curriculum trends. *Prospects, 54*(3), 645–660.

Karumbaiah, S., Borchers, C., Shou, T., Falhs, A.-C., Liu, P., Nagashima, T., Rummel, N., & Aleven, V. (2023). A spatiotemporal analysis of teacher practices in supporting student learning and engagement in an AI-enabled classroom. In *International Conference on Artificial Intelligence in Education* (pp. 450–462). Springer Nature Switzerland.

Khabarov, V., Volegzhanina, I., & Volegzhanina, E. (2023). Ontology-based AI mentor for training future "Digital Railway" engineers. In *International Scientific Conference Fundamental and Applied Scientific Research in the Development of Agriculture in the Far East* (pp. 31–42). Springer Nature Switzerland.

Kim, W.-H., & Kim, J.-H. (2020). Individualized AI tutor based on developmental learning networks. *IEEE Access, 8*, 27927–27937.

Kloos, C. D., Alario-Hoyos, C., Estévez-Ayres, I., Callejo Pinardo, P., Hombrados-Herrera, M. A., Muñoz-Merino, P. J., Moreno-Marcos, P. M., Muñoz-Organero, M., & Ibáñez, M. B. (2024). How can generative AI support education? In *2024 IEEE Global Engineering Education Conference (EDUCON)* (pp. 1–7). IEEE.

Lawal, A. (2024). *STEM teacher shortage in American schools: A case study exploring the recruitment of STEM majors into teacher education programs* (Doctoral dissertation). University of North Dakota.

Létourneau, A., Deslandes Martineau, M., Charland, P., Karran, J. A., Boasen, J., & Léger, P. M. (2025). A systematic review of AI-driven intelligent tutoring systems (ITS) in K–12 education. *npj Science of Learning, 10*(1), 29.

Llanos, R., Gonzales, G., & Morzan, J. (2024). Enhancing student success through AI integration: A study on the implementation of a virtual assistant in higher education courses. In *2024 IEEE 4th International Conference on Advanced Learning Technologies on Education & Research (ICALTER)* (pp. 1–4). IEEE.

Lee, S. J., & Kwon, K. (2024). A systematic review of AI education in K-12 classrooms from 2018 to 2023: Topics, strategies, and learning outcomes. *Computers and Education: Artificial Intelligence, 6*, 100211.

Looi, C.-K., & Jia, F. (2025). Personalization capabilities of current technology chatbots in a learning environment: An analysis of student–tutor bot interactions. *Education and Information Technologies*, 1–31.

Lui, R. W. C., Bai, H., Zhang, A. W. Y., & Chu, E. T. H. (2024). GPTutor: A generative AI-powered intelligent tutoring system to support interactive learning with knowledge-grounded question answering. In *2024 International Conference on Advances in Electrical Engineering and Computer Applications (AEECA)* (pp. 702–707). IEEE.

Madhuri, J. N., Fatma, G., Kumari, A., Pathak, P., & Mohamed Faizal, M. (2024). Creating an AI English tutor with personalized content and dynamic lessons. In *2024 International Conference on Artificial Intelligence and Quantum Computation-Based Sensor Application (ICAIQSA)* (pp. 1–6). IEEE.

Mailach, A., Gorgosch, D., Siegmund, N., & Siegmund, J. (2025). "Ok Pal, we have to code that now": Interaction patterns of programming beginners with a conversational chatbot. *Empirical Software Engineering, 30*(1), 34.

Maity, S., & Deroy, A. (2024). Generative AI and its impact on personalized intelligent tutoring systems. *arXiv preprint*, arXiv:2410.10650.

Makharia, R., Kim, Y. C., Jo, S. B., Kim, M. A., Jain, A., Agarwal, P., Srivastava, A., Agarwal, A. V., & Agarwal, P. (2024). AI tutor enhanced with prompt engineering and deep knowledge tracing. In *2024 IEEE International Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation (IATMSI)* (Vol. 2, pp. 1–6). IEEE.

Maxwell, J. A. (2012). The importance of qualitative research for causal explanation in education. *Qualitative Inquiry, 18*(8), 655–661.

Mousavinasab, E., Zarifsanaiey, N., Niakan Kalhori, S. R., Rakhshan, M., Keikha, L., & Ghazi Saeedi, M. (2021). Intelligent tutoring systems: A systematic review of characteristics, applications, and evaluation methods. *Interactive Learning Environments, 29*(1), 142–163.

Nwana, H. S. (1990). Intelligent tutoring systems: An overview. *Artificial Intelligence Review, 4*(4), 251–277.

Paudyal, P., Banerjee, A., & Gupta, S. (2020). On evaluating the effects of feedback for sign language learning using explainable AI. In *Companion Proceedings of the 25th International Conference on Intelligent User Interfaces* (pp. 83–84).

Peñalvo, F. J. G., & Vázquez Ingelmo, A. (2023). What do we mean by GenAI? A systematic mapping of the evolution, trends, and techniques involved in Generative AI. *IJIMAI, 8*(4), 7–16.

Peterson, R. A., & Merunka, D. R. (2014). Convenience samples of college students and research reproducibility. *Journal of Business Research, 67*(5), 1035–1041.

Roumeliotis, K. I., & Tselikas, N. D. (2023). ChatGPT and Open-AI models: A preliminary review. *Future Internet, 15*(6), 192.

Sarshartehrani, F., Mohammadrezaei, E., Behravan, M., & Gracanin, D. (2024). Enhancing e-learning experience through embodied AI tutors in immersive virtual environments: A multifaceted approach for personalized educational adaptation. In *International Conference on Human-Computer Interaction* (pp. 272–287). Springer Nature Switzerland.

Shahri, H., Emad, M., Ibrahim, N., Rais, R. N. B., & Al-Fayoumi, Y. (2024). Elevating education through AI tutor: Utilizing GPT-4 for personalized learning. In *2024 15th Annual Undergraduate Research Conference on Applied Computing (URC)* (pp. 1–5). IEEE.

Soliman, H., Kravcik, M., Neumann, A. T., Yin, Y., Pengel, N., & Haag, M. (2024). Scalable mentoring support with a large language model chatbot. In *European Conference on Technology Enhanced Learning* (pp. 260–266). Springer Nature Switzerland.

Soudani, H., Kanoulas, E., & Hasibi, F. (2024). Fine tuning vs. retrieval augmented generation for less popular knowledge. In *Proceedings of the 2024 Annual International ACM SIGIR Conference on Research and Development in Information Retrieval in the Asia Pacific Region* (pp. 12–22).

Vijay, V. C., Lees, M., & Vakaj, E. (2020). Introducing knowledge-based augmented reality environment in engineering learning: A comparative study. In *2020 IEEE Learning With MOOCS (LWMOOCS)* (pp. 131–143). IEEE.

Wu, T., He, S., Liu, J., Sun, S., Liu, K., Han, Q.-L., & Tang, Y. (2023). A brief overview of ChatGPT: The history, status quo and potential future development. *IEEE/CAA Journal of Automatica Sinica, 10*(5), 1122–1136.

Yang, J., Cooper-Stachowsky, M., & Kamal, Z. H. (2024). On implementing an effective intelligent tutor and its impact on teaching and learning experiences. In *International Conference on Artificial Intelligence in Education Technology* (pp. 171–190). Springer Nature Singapore.

Yang, K. B., Nagashima, T., Yao, J., Williams, J. J., Holstein, K., & Aleven, V. (2021). Can crowds customize instructional materials with minimal expert guidance? Exploring teacher-guided crowdsourcing for improving hints in an AI-based tutor. *Proceedings of the ACM on Human-Computer Interaction, 5*(CSCW1), 1–24.