

www.ijte.net

Benefits, Challenges, and Methods of Artificial Intelligence (AI) Chatbots in Education: A Systematic Literature Review

Şahin Gökçearslan
Dıstance Educatıon R & D Center, Turkey

Cansel Tosun
Gazi University, Turkey

Zeynep Gizem Erdemir
Gazi University, Turkey

To cite this article:

Gokcearslan, S., Tosun, C., & Erdemir, Z.G. (2024). Benefits, challenges, and methods of Artificial Intelligence (AI) chatbots in education: A systematic literature review. *International Journal of Technology in Education (IJTE), 7*(1), 19-39. https://doi.org/10.46328/ijte.600

The International Journal of Technology in Education (IJTE) is a peer-reviewed scholarly online journal. This article may be used for research, teaching, and private study purposes. Authors alone are responsible for the contents of their articles. The journal owns the copyright of the articles. The publisher shall not be liable for any loss, actions, claims, proceedings, demand, or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of the research material. All authors are requested to disclose any actual or potential conflict of interest including any financial, personal or other relationships with other people or organizations regarding the submitted work.

This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

International Journal of Technology in Education
2024, Vol. 7, No. 1, 19-39 https://doi.org/10.46328/ijte.600

Benefits, Challenges, and Methods of Artificial Intelligence (AI) Chatbots in Education: A Systematic Literature Review

Şahin Gökçearslan, Cansel Tosun, Zeynep Gizem Erdemir

Article Info

Article History

Received:
09 July 2023
Accepted:
21 December 2023

Keywords

AI chatbots
Education
Systematic literature review
Method

Abstract

In many fields, AI chatbots continue to be popular with new tools and attract the attention of universities, K12 schools, educational organizations, and researchers. The aim of this research is to review the research on AI chatbots by restricting it to the category of education and to examine this research from a methodological point of view. Therefore, we performed a systematic literature review with a sample of 37 SSCI articles published in the educational context. Within the scope of the selected studies, the advantages and disadvantages of AI chatbots in education for students and educators, as well as the types of chatbots used, year, keywords, and method were analyzed. According to the research results, increased motivation to learn and language skill development are advantages for students, while cost-effectiveness and reduced workload are advantages for educators. Limited interaction, misleading answers for learners, originality, and plagiarism are the most common disadvantages for educators. The study also includes research results and recommendations related to the methodological review.

Introduction

While artificial intelligence is a technology that can be traced back to the first computers, it has brought artificial intelligence-supported (AI) chatbots to its users, which have gained great popularity worldwide with language models called generative artificial intelligence and are growing rapidly. Chatbots are being utilized in various fields and are referred to as 'conversational agents', 'conversational tutors', or simply just 'bots' (Pérez et al., 2020). One of the AI-supported chatbots is ChatGPT. ChatGPT, an affiliate of some of the online services such as Instagram, Twitter and Netflix took 2.5 months to 3.5 years to reach one million users. Chat GPT met its users in November 2022 and this period is 5 days (Buchholz, 2023). ChatGPT, which is rapidly gaining popularity and attracting experimenters around the world, reached 100 million users just two months after its launch (Sabzalieva & Valentini, 2023). As AI develops, AI-enabled chatbots used for various needs and purposes, including, marketing, customer service, technical support and education (Smutny & Schreiberova, 2020). According to the 2023 Educause Horizon Report Teaching and Learning Trends, AI in 2021, AI for learning Analytic in 2022 and AI-anabled applications for personal and predictive learning and generative AI is key technologies in education in 2023.

According to the report, the revolutionary features of generative AI extend beyond the learning environment.

People may have the opportunity to be freed from mundane tasks in classroom and administrative work (Educause, 2023). ChatGPT represents the current state of artificial intelligence, processing information in an efficient, systematic and informative way and is said to reform education. Different AI chatbots are available for training and can be used to support different learning outcomes. AI Chatbots are intelligent systems developed in the one hand, the use of AI chatbots in education can bring about many benefits, including increased student engagement, collaboration, and access, on the other hand, these tools can cause a range of new issues and concerns, particularly to scientific integrity and academic plagiarism (Cotton et al., 2023).

The effectiveness of AI chatbots depends on their ability to understand text and produce human-like text. The comprehension capability of this tool is based on the data and algorithms it encompasses. Since AI chatbots, so-called generative AI technologies, have the potential to influence teaching and learning processes and methods in an unconventional way and require educators to develop new ways of thinking and keep up with the transformation, it is emphasized that the potential benefits of the technology should be used carefully and cautiously (Bozkurt, 2023). As ChatGPT becomes more widely used in educational settings, its applications need to be managed responsibly and ethically (Mhlanga, 2023).

Although there is still confusion about the use of AI chatbots in education, various organizations are closely following developments in this field. In 2023, UNESCO developed a quick guide to AI applications in higher education with a specific focus on ChatGPT (Sabzalieva & Valentini, 2023). The organization has also published a research report that includes ethical values as a guideline for AI curricula in the K12 learning category, with a special focus on ChatGPT (UNESCO, 2022). Many AI-powered chatbots exist and are being developed with different techniques. The first version of ChatGPT, one of them, started with the first-generation generative pre-trained transducer (GPT) language model, which is an early evolution of Natural Language Processing (NLP) algorithms (Zhang & Li, 2021).

Training techniques such as reinforcement or transfer learning are utilized to improve the performance of GPT and NLP algorithms and chatbots (OpenAI, 2023). ChatGPT encourages follow-up questions through a continuous dialog, providing a different experience from search engines. Search engines do not store the evolving history of the answer, they just list it. ChatGPT responds to the challenges posed by the questioner by providing follow-up questions that develop and expand the answers (Rospigliosi, 2023). The GPT-3 model is a language model that uses deep learning to produce human-like text.

This language has been trained on trillions of text documents (Brown et al., 2020). OpenAI released GPT-4 model in March 2023, which will have trillions of parameters compared to its predecessor's millions. Additionally, this new version will be capable of generating images and videos along with text (Loafers, 2023). While OpenAI was introducing GPT models to the user, many AI chatbots were derived using this model. In 2018, apart from OpenAI, Google Research developed BERT (Bidirectional Encoder Representations from Transformers) (Devlin et al., 2018). In 2019, GoogleAI released XLNet model (Yang et al., 2019). With other models developed, many AI chatbots point to the transformation in education as a technology tool with different features to support learning activities.

The use of AI chatbots in education has been associated with advantages such as personalized learning, speed of access to information, instant feedback. On the other hand, the risk of reducing students' cognitive skills is also expressed. In addition, it is important to develop students' digital literacy skills to effectively integrate AI chatbots into learning processes (Farrokhnia et al., 2023). AI chatbots can be used for purposes such as analyzing and evaluating students' learning abilities (Durall & Kapros, 2020), management and teacher assistant (Okonkwo & Ade-Ibijola, 2021).

Systematic review of international research is a guiding research area to determine which variables are focused on and to identify strengths as well as to investigate them. Systematic analysis of research on technological developments and the use of technology in education sheds light on future studies in terms of providing an overview of the field/subject area (Hsu et al., 2013). While research on the effects of new technologies in education continues, systematic analyses of many current technologies are included. Augmented reality (Akçayır & Akçayır, 2017), artificial intelligence (Chen et al., 2020), artificial intelligence in online and distance education (Dogan et al., 2023), Internet of things (IoT) (Kassab et al., 2020), virtual reality (Kavanagh et al., 2017), blockchain applications (Delgado-von-Eitzen et al., 2021) are among them. Although the interest of researchers in the use of AI chatbots in education and other fields is increasing, it is considered important to conduct systematic review studies for future research and to eliminate the confusion created by the change with AI chatbot technologies and to eliminate the research gap in the field (Dwivedi et al., 2023). In literature, a systematic literature review of 53 articles on chatbots and conversational agents in various international databases was found (Okonkwo & Ade-Ibijola, 2021). There is limited systematic literature review specifically on artificial intelligence chatbots. In this context, the aim of the current study is to conduct a systematic review of AI chatbots in the educational category of research and to examine the research from a methodological perspective.

Interest in the use of AI chatbots for teaching and learning purposes is increasing day by day (Smutny & Schreiberova, 2020). When the literature is examined, it is seen that many systematic literature reviews have been conducted with current developments in AI. In these systematic reviews, fields such as technology (Tang et al., 2023), architecture (Yiğitcanlar et al., 2020), health (Sapçı & Sapçı, 2020), mathematics (Hwang & Tu, 2021), educational sciences (Zawacki-Richter et al., 2019) were included. Systematic literature reviews in the field of educational sciences include the use of AI in language education (Liang et al., 2023), artificial intelligence in teacher education (Salas-Pilco et al., 2020), trends in AI-supported e-learning (Tang et al., 2023), and the use of artificial intelligence for assessment (González-Calatayud et al., 2021).

When the literature is examined, in a study examining the use of AI chatbots in the field of education, the types of AI chatbots were categorized and the purposes for which they are used in teaching were examined (Perez et al., 2020). In another study, the usage areas and pedagogical roles of AI chatbots were investigated (Wollny et al., 2021). In another systematic literature review, variables such as design principles, interaction styles, etc. related to chatbots were included (Kuhail et al., 2023). In a systematic literature review conducted by Okonkwo and Abejide-Ade (2021), the benefits and challenges of using chatbot technologies in education as well as their use for various purposes in education were examined. In this study, the advantages and disadvantages of AI-supported chatbots for students and educators are presented. In addition, the types, years, keywords and research methods

of chatbots frequently used in education were analyzed."

Examining the methodological and general trends of the studies examining the use of AI chatbots in education in depth with a holistic perspective will be useful in identifying the research gap that will guide future research. In this context, studies published in SSCI indexed journals between 2020-2023 in the category of education related to AI chatbots were analyzed and the following questions were sought to be answered with this study:

- What are the advantages of AI chatbots in educational environments?
- What are the disadvantages of AI chatbots in educational environments?
- What is the distribution of publications on the use of AI chatbots for educational purposes in SSCI indexed journals by years?
- What is the keyword distribution of the publications on the use of AI chatbots for educational purposes in SSCI indexed journals?
- What is the distribution of AI chatbots used in publications on the use of artificial intelligence chat robots for educational purposes in SSCI indexed journals?
- What is the method distribution of publications on the use of AI chatbots for educational purposes in SSCI indexed journals?
- What is the data analysis method distribution of the publications on the use of AI chatbots for educational purposes in SSCI indexed journals?
- What is the distribution of sample types of publications on the use of AI chatbots for educational purposes in SSCI indexed journals?
- What is the sample size distribution of the publications on the use of AI chatbots for educational purposes in SSCI indexed journals?
- What is the distribution of data collection tools in publications on the use of AI chatbots for educational purposes in SSCI indexed journals? "

Method

The Manuscript Selection Process

In systematic review studies, searches on the subject are made in various databases. This review includes a systematic analysis of their publications in journals in the Web of Science database indexed in Social Science Citation Index (SSCI). The SSCI index was chosen because it includes publications in top-tier journals (Gursoy & Sandstrom, 2016) in the field of social sciences in Web of Science. It is also possible to restrict research reports in the SSCI database with labels such as subject and research area (Luor et al., 2008). In this study, the limitations stated in the table below were made.

In order to reach all publications related to chatbots, the variables Chat* and artificial intelligence were selected in the TS (Topic) category. TS searches for topic terms in the following fields within a record include title, abstract, author keywords, and Keywords Plus. No time frame is specified when conducting the search. The last search was conducted on 12 April 2023.

Inclusion and exclusion criteria are given Table 1.

Table 1. Inclusion and Exclusion Criteria

Inclusion criteria	Exclusion criteria
• chat* and artificial intelligence in TS • The article should be about the use of chatbots for educational purposes • Article or Early Access or Review Article	• Editorials are excluded. • Indexes other than SSCI • Proceeding Paper is excluded

Query of current research: Results for TS= “chat*”and TS= “artificial intelligence” and Article or Early Access or Review Article (Document Types) and Education Educational Research or Education Scientific Disciplines (Web of Science Categories) and Social Sciences Citation Index (SSCI) (Web of Science Index) and Proceeding Paper (Exclude – Document Types) .The search for "chat* and artificial intelligence in TS" yielded 821 articles. When education discipline and SSCI articles were included and Proceeding paper is excluded, 37 articles were found.

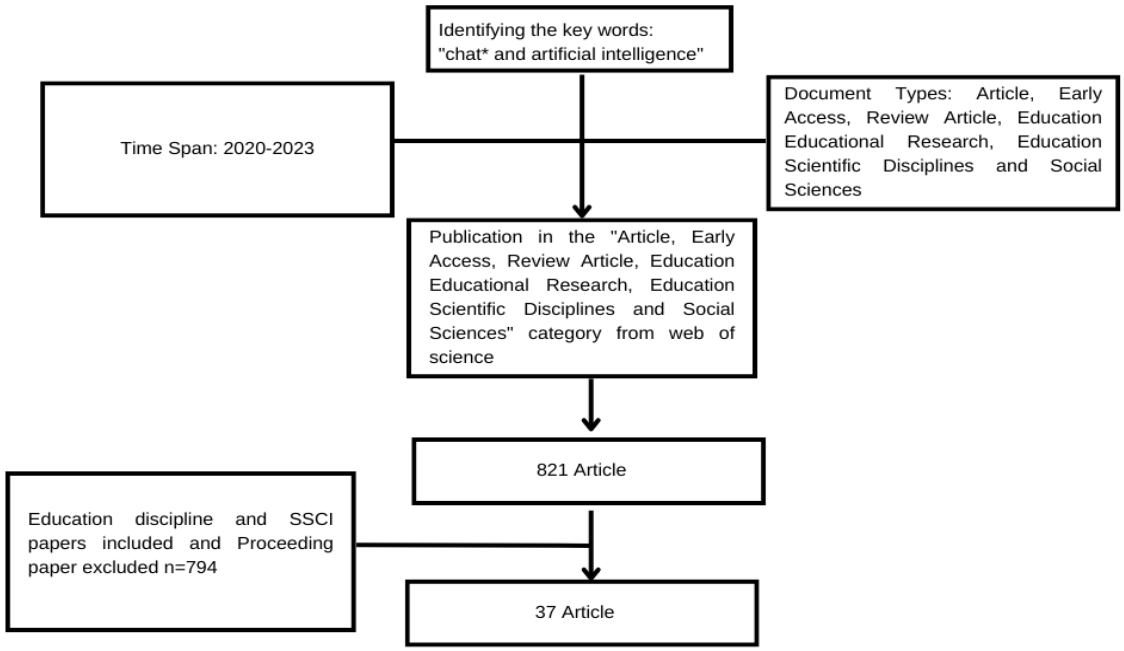

Figure 1. Selection Diagram of Articles

The Data Coding and Analysis Processes

According to the content of the 37 articles obtained from the Web of Science database, the researchers checked whether they were directly related to "artificial intelligence and chatbots". The pre-examined articles were

transferred from the database to M.S. Office Excel. The classification in the research method is based on Sözbilir, Kutu and Yaşar (2012). The article classification form consists of six sections: article title, subject, method, data collection tools, sample and data analysis methods. Based on the Article Classification Form, the method, data collection tools, sample, and data analysis method categories of the article were analyzed. If the feature being analyzed is not listed in the classification, it is defined as "other" in this study; if the feature is not mentioned in the article, it is defined as "unspecified". The most used AI Chatbot technologies, common keywords, and distribution by years in the articles were determined by content analysis method. The year is the publication date of the article in the specified journal. To determine the advantages and disadvantages section, the main themes were identified by reading all findings, conclusions and discussion sections of 37 articles. The data obtained from the articles were analyzed with the content analysis technique in the next stage. Content analysis was preferred to meet the objectives of this research as it allows for the examination of the connections between the elements. Tech's 8 steps were used to open code the data (Tesch, 2013). Frequency/percentage values were used to analyze the data.

The data were analyzed in three stages. First, the publications were randomly shared among the researchers and classified. The classified data were re-examined and undecided situations were resolved by consulting an expert. Finally, 19% (n=7) of the publication data entered by the researchers were analyzed by other researchers. The agreement values between the coders were examined by determining the disagreements and differences. Similar codes are called "Agreement" and dissimilar codes are called "Disagreement" and the formula for coder reliability is Percent Agreement = Agreement / (Agreement + Disagreement) * 100 (Miles & Huberman, 1994, p.64). As a result of the analysis, it was calculated that the agreement between the coders was at an acceptable (at least 70%) level (93.75%) (Miles & Huberman, 1994, p.64).

Findings

RQ1: What Are the Advantages of AI Chatbots in Educational Environments as Stated in Studies Published in SSCI Indexed Journals?

The findings regarding the advantages of AI chatbots in educational environments are given in Table 2.

Table 2. Advantages of AI Chatbots in Education

Themes	Sub-categories	f	Sample research
Advantages for the student	Increased motivation to learn	15	Chang et al., 2022
	Language skill development	14	Vazquez-Cano et al., 2021
	Improved learning performance	10	Deveci Topal et al., 2021
	Personalized and adaptive learning environments	10	Ji et al., 2023

Themes	Sub-categories	f	Sample research
	Instant feedback	8	Cotton et al., 2023
	Ease of access	7	Liu et al., 2022
	Self-regulated learning	5	Han et al., 2022
	Active participation	2	Xia et al., 2023
Benefits for the educator	Economy in time	2	Ji et al., 2023
	Workload reduction	2	Cooper, 2023
	Providing training resources	1	Cotton et al., 2023

Among the advantages for the learner, increased motivation to learn, language skill development, increased learning performance, personalized and adaptive learning environments are the most common themes. In addition, immediate feedback, ease of access, self-regulated learning and active participation were also common themes. Educator-oriented themes include affordability, reducing workload and providing training resources.

RQ2: What Are the Disadvantages of AI Chatbots in Educational Environments as Stated in Studies Published in SSCI Indexed Journals?

The findings regarding the disadvantages of AI chatbots in educational environments are given in Table 3.

Table 3. Disadvantages of AI Chatbots in Education

Themes	Sub-categories	f	Sample research
Disadvantages for the student	Limitation of interaction	7	Xu et al., 2021
	Misleading answers	6	Cooper, 2023
	Limitation of personalized feedback	6	Guo et al., 2022
	Failure to understand complex expressions	5	Xia et al., 2023
	Outdated responses	3	Essel et al., 2022
	Decreased communication skills	1	Godwin-Jones, 2023
	Mechanics of voice and intonation	1	Yang et al., 2022
	Difficulty getting used to technology	1	Ji et al., 2023

Themes	Sub-categories	f	Sample research
	Data protection and privacy	1	Moldt et al., 2023
Disadvantages for the educator	Originality and plagiarism	4	Cotton et al., 2023
	Failure to determine the level of readiness	3	Hsu et al., 2023
	The challenge of developing AI-powered applications	1	Chen et al., 2023

In Table 3, the disadvantages of AI chatbots in student-oriented education include limited interaction, misleading answers, limited personalized feedback, not understanding complex expressions, outdated answers, decreased communication skills, mechanical voice intonation, difficulty in getting used to technology, data protection and privacy. Disadvantages for the educator include originality and plagiarism, inability to determine the level of readiness, and difficulty in developing AI-supported applications.

RQ3: What is the Distribution over Time of Studies Published in SSCI Indexed Journals that Examine the Use of Artificial Intelligence Chatbots for Educational Purposes?

The distribution of publications on AI chat tool is given in Figure 2.

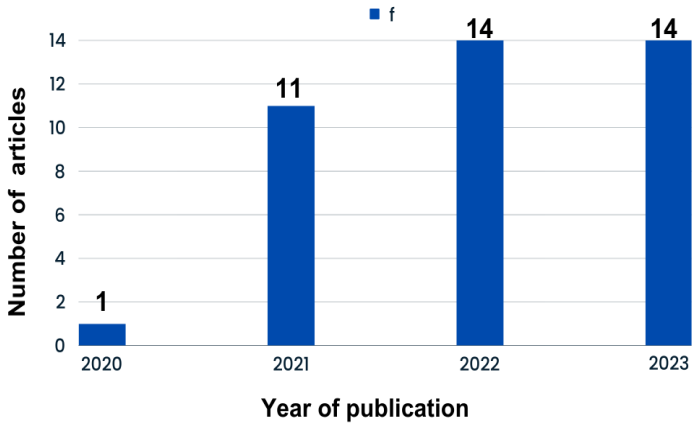

Figure 2. Distribution of Publications on AI Chatbot by Years

According to Figure 2, there was only one published study on AI chatbot in 2020, the number of studies increased in 2021 (n=11) and continued to increase in 2022 (n=14) and 2023 (n=14).

RQ4: What Is the Distribution of Keywords in Studies Published in SSCI Indexed Journals That Examine the use of Artificial Intelligence Chatbots for Educational Purposes?

The distribution of keywords in publications related to AI chat tool is given in Figure 3.

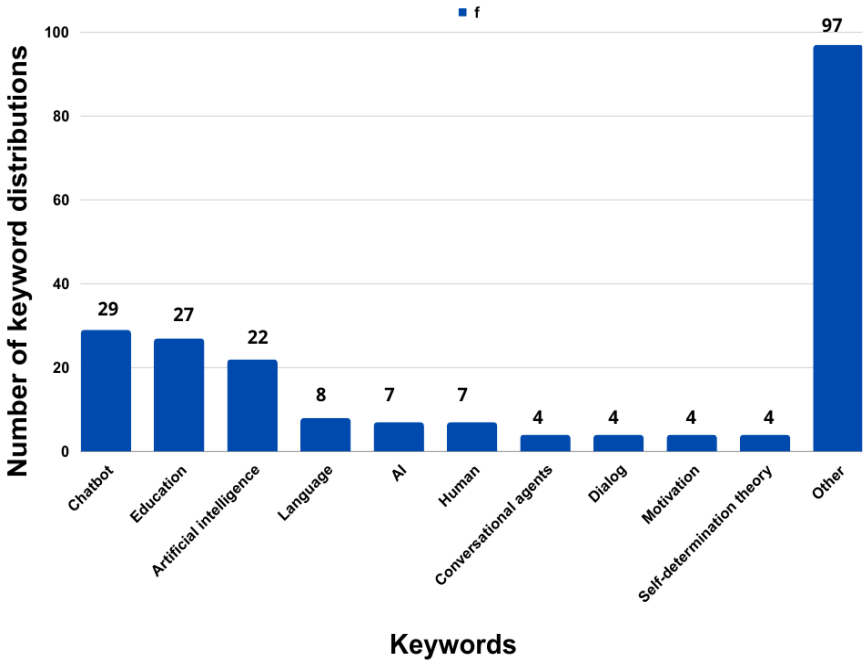

Figure 3. Distribution of Keywords in Publications Related to AI Chatbot

The distribution of keywords in the publications related to AI chatbot is given in Figure 3. According to Figure 3, it is determined that chatbot (n=29), education (n=27), artificial intelligence (n=22), language (n=8), AI (n=7), Human (n=7) are the most used keywords in published publications.

RQ5: What Is the Distribution of AI Chatbots Commonly Used in Research Published in SSCI Indexed Journals That Examine the Use of Artificial Intelligence Chatbots for Educational Purposes?

The distribution of commonly used AI chatbots in publications is given in Table 4.

Table 4. AI Chatbots Commonly used in Publications

	f	%
Other AI Chatbots	22	59.5
Unspecified	10	27.0
ChatGPT	3	8.1
Ellie	2	5.4
Total	37	100.0

The AI chatbots commonly used in the studies are shown in Table 4. According to Table 4, it is determined that ChatGPT 8.1% (f=3), Ellie 5.4% (f=2), other chatbots 59.5% (f=22) are commonly used AI chatbots.

RQ6: What Is the Distribution of Method in Research Published in SSCI Indexed Journals that Examine the Use of Artificial Intelligence Chatbots for Educational Purposes?

The distribution of method in publications is given in Table 5.

Table 5. Distribution of the Methods in Publications on AI Chatbots

	Research Design	Research Method	f	%
Quantitative	Experimental	Quasi-experimental	15	40.5
		Weak experimental	1	2.7
		Sub-total	16	43.2
	Non-experimental	Comparison	3	8.1
		Sub-total	3	8.1
Qualitative	Interactive	Other	2	5.4
		Phenomenology	1	2.7
		Sub-total	3	8.1
	Non-interactive	Review	4	10.8
		Systematic review	4	10.8
		Other	1	2.7
		Sub-total	9	24.3
Mixed design	Mixed	Explanatory (Quantitative - Qualitative)	4	10.8
		Triangulation (Quantitative + Qualitative)	2	5.4
		Sub-total	6	16.2
		Total	37	100.0

The distribution of the preferred method in the publications on AI chatbot is given in Table 5. It was analyzed that among the quantitative research methods, the Quasi-Experimental research method (40.5%), which is an experimental design, was preferred the most. In qualitative research methods, review (10.8%) and systematic review (10.8%) research methods, which are non-interactive research designs, were used the most. In mixed research methods, it was determined that they used Explanatory (Quantitative - Qualitative) (quantitative-

qualitative) (10.8%). In quantitative research methods, comparative (8.1%), which is a non- experimental design, was used. In mixed research methods, mixed design Triangulation (Quantitative + Qualitative) (5.4%) research method was used. In qualitative research methods, Interactive design Other (5.4%) research methods were used. Experimental research design from quantitative research methods quasi-experimental design (2.7%) research method was used. In qualitative research, case study (2.7%) research method with interactive research design was used. Among qualitative research methods, non-interactive research design, other (2.7%) research method was used.

RQ7: What Is the Distribution of Data Analysis Method in Research Published in SSCI Indexed Journals that Examine the use of Artificial Intelligence Chatbots for Educational Purposes?

The distribution of data analysis method in publications is given in Table 6.

Table 6. Findings on Data Analysis Method in Publications on AI Chatbots in Education

Main Category	Subcategory	f	%
Descriptive	Frequency/Percentage	18	43.9
	Mean/Standard deviation	14	34.1
	Graphical representation	6	14.6
	Other	3	7.3
	Subtotal	41	100.0
Predictive	T-test	6	25.0
	ANOVA/ANCOVA	4	16.7
	Correlation	4	16.7
	Regression	3	12.5
	Other	3	12.5
	Factor analysis	2	8.3
	Non-parametric tests	2	8.3
	Subtotal	24	100.0
Qualitative	Qualitative descriptive analysis	13	54.2
	Other	7	29.2
	Content analysis	4	16.7
	Subtotal	24	100.0

When Table 6 is examined, it is seen that frequency/percentage method is used most frequently with a rate of

43.9% (f=18) among quantitative descriptive analysis methods. This is followed by mean/standard deviation with 34.1% (f=14), graphical representation with 14.6% (f=6) and other with 7.3% (f=3). Among quantitative predictive methods, t-test was the most preferred with 25.0% (f=6), followed by correlation with 16.7% (f=4) and ANOVA/ANCOVA with 16.7% (f=4), regression with 12.5% (f=3) and other with 12.5% (f=3), factor analysis with 8.3% (f=2) and non-parametric tests with 8.3% (f=2). The most preferred qualitative data analysis method was descriptive analysis 54.2% (f=13), followed by other 29.2% (f=7) and content analysis 16.7% (f=4).

RQ8: What Is the Distribution of Sample Types of Frequently Studied Research in SSCI Indexed Journals that Examine the use of Artificial Intelligence Chatbots for Educational Purposes?

The distribution of sample types frequently studied in publications is given in Table 7.

Table 7. Sample Types Frequently Studied in Publications on AI Chatbots in Education*

Sample Types Frequently Examined in Research*	f	%
Undergraduate	14	35.9
Other	10	25.6
Secondary (9-12)	7	17.9
Primary (6-8)	4	10.3
Preschool	1	2.6
Primary (1-5)	1	2.6
Graduate	1	2.6
Teacher	1	2.6
Total	39	100.0

* There are researches with more than one sample type.

As seen in Table 7, it was determined that the most frequently studied sample types in publications on AI chatbot in education are undergraduate students (35.9%) and other (25.6%) sample types. These were followed by secondary school students (17.9%), 6-8 primary school students (10.3%), preschool students (2.6%), 1-5 primary school students (2.6%), graduate students (2.6%) and teachers (2.6%).

RQ9: What Is the Distribution of Sample Size Frequently Studied Research in SSCI Indexed Journals That Examine the use of Artificial Intelligence Chatbots for Educational Purposes?

The distribution of sample size frequently studied in publications is given in Figure 4.

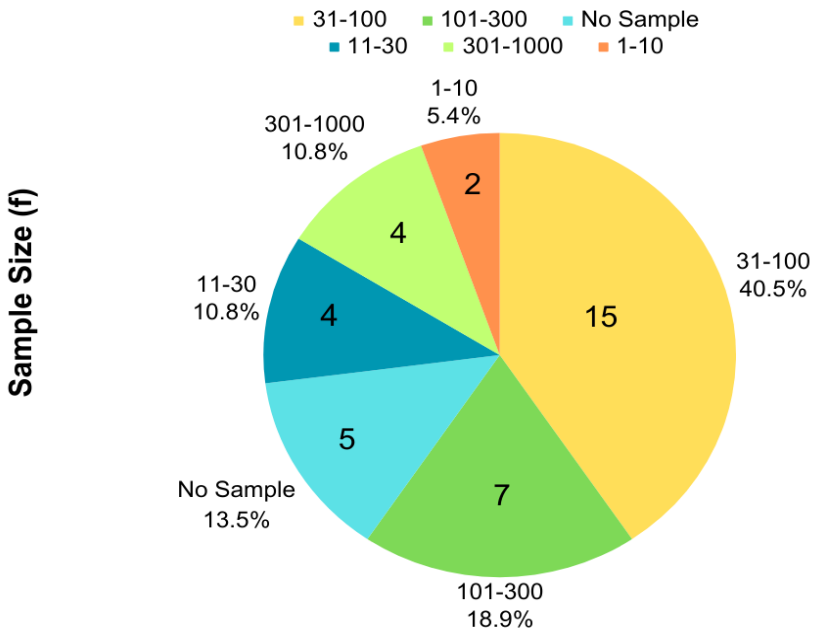

Figure 4. Sample Size Frequently Studied in Publications on AI Chatbot in Education

When Figure 4. is examined, it is seen that 40.5% (f=15) sample size is mostly used in the 31-100 range. This finding is followed by 18.9% (f=7) in the 101-300 range. In addition, 13.5% (f=5) sample was not used. Then, 10.8% (f=4) and 10.8% (f=4) sample sizes were used in the range of 11-30, 301-1000, and 5.4% (f=2) in the range of 1-10.

RQ11: What Is the Distribution of Preferred Data Collection Tools in SSCI Indexed Journals Examining the Use of Artificial Intelligence Chatbots for Educational Purposes?

The distribution of s preferred data collection tools in SSCI indexed journals is given in Table 8.

Table 8. Frequently Preferred Data Collection Tools in Publications on AI Chatbot in Education*

Data collection tools	f	%
Interview	12	23.5
Perception-interest-attitude-ability tests	10	19.6
Achievement Tests	10	19.6
Questionnaire	8	15.7
Documents	5	9.8
Other	4	7.8

Data collection tools	f	%
Alternative assessment tools (Diagnostic test, concept maps, portfolio Etc.)	1	2.0
Observation	1	2.0
Total	51	100.0

* More than one option is checked

As seen in Table 8, it is seen that interviews are the most frequently used data collection tools in studies on AI chatbots in education with 23.5% (f=12). This tool was followed by perception-interest-attitude-ability etc. tests, 19.6% (f=10) and achievement tests 19.6% (f=10), questionnaire 15.7% (f=8), documents 9.8% (f=5), other 7.8% (f=4), alternative assessment tools (diagnostic test, concept maps, portfolio etc.) 2.0% (f=1) and observation 2.0% (f=1).

Discussion

In terms of the advantages of AI chatbots in education, among the advantages of AI chatbots for students, increased motivation to learn, language skill development, and increased learning performance are the most frequently encountered themes in the research. The advantages of AI chatbots for the educator include cost-effectiveness, reducing workload, and providing educational resources. In systematic literature reviews on the integration of technology-supported tools into the educational environment, student engagement and motivation come to the fore (Banihashem et al., 2023; Dehghanzadeh et al., 2023). The integration of AI chatbots into education can be examined in a similar framework.

Among the disadvantages of AI chatbots in education for students, limited interaction, misleading answers, and limited personalized feedback are the most common ones. The disadvantages of AI chatbots for the trainer include originality and plagiarism, inability to determine the level of readiness, and difficulty in developing AI-supported applications. AI chatbots have started to be developed and steps towards solving ethical problems over time have begun to be seen. Goswami (2023) stated that OpenAI announced that user data will not be used in AI chatbots. In the first months of 2023, similarity and plagiarism problems can now be prevented by determining the development status of that content with AI based on GPT algorithms in the tools of companies that detect similarity. All these processes pave the way for educators to reconsider their approaches to education planning, implementation, and evaluation.

In 2020, there was one published research on AI chatbot, while more than 10 publications were found in 2021, 2022 and 2023, although it is only in the 3rd month. According to Paek and Kim (2021), the number of articles published on general AI and education increases exponentially every year according to the research limited to SCI-EXPANDED, SSCI, A&HCI, and ESCI indexes and the education category in the web of science. Since 2010, the process has been called the "third AI boom" (Rose et al., 2016).

It was determined that chatbot, education, artificial intelligence, language, AI, human were the most frequently used keywords in the publications. When it was analyzed which AI chatbot was frequently used in the studies, ChatGPT (n=3), Ellie (n=2) were the most frequently used AI chatbot tool, while other chatbot types were used to a great extent (60%). AI chatbots whose type is not specified in the publications are slightly more than 1 in 4. Baidoo-Anu and Owusu Ansah (2023) highlight that generative AI, as used by ChatGPT, has the potential to revolutionize teaching and learning methods by providing students with personalized instruction, feedback and automatic essay grading (Chen et al., 2020; Kim et al., 2019). Other tools are also being developed to address different learning needs, building on the potential of ChatGPT.

When the published studies on AI chatbot are examined, as research methods, quantitative, mixed and qualitative methods were used most frequently in the studies, respectively. Quasi-experimental design among experimental designs, review and systematic review in qualitative research, and explanatory (Quantitative + Qualitative) design in mixed research method are the most frequently used designs. In a study examining research on the use of AI technologies in education, bibliometric analysis, categorical meta-trend analysis and inductive content analysis were conducted. According to the results of a similar study, quantitative + qualitative mixed design method was frequently used (Zhang & Aslan, 2021).

In the publications on AI chatbot in education, frequency/percentage tables (1) were the most common data analysis method among quantitative descriptive analysis methods, followed by mean/standard deviation (2), graphical representation (3) and other methods. Among quantitative predictive methods, t-test (1) was followed by correlation (2), ANOVA/ANCOVA (3), regression (4) and other analyses. The most preferred qualitative data analysis method was qualitative descriptive analysis (1), followed by other analyses (2) and content analysis (3). In a previous study on the use of chatbots in education, SSCI index publications were analyzed in the WOS database. According to a study, quantitative methods such as ANOVA, descriptive statistics, t-test and correlation analysis are frequently preferred (Hwang & Chang, 2021).

Undergraduate students (1), other sample types (2), and secondary education students (3) were the most frequently studied sample types in publications on AI chatbot in education. In previous systematic reviews on the use of chatbots in education, it was found that studies on artificial intelligence in higher education were more common as a sample type (Zhang & Aslan, 2021; Nee et al., 2023). The sample size frequently studied in publications is 31-100 range (1), followed by 101-300 (2), 11-30 (3). In addition to interviews and interviews (1), perception-interest-attitude-ability tests (2) and achievement tests (3), alternative assessment tools (diagnostic tests, concept maps, portfolios, etc.) were also used. In a study conducted by Hwang and Chang (2021) on the opportunities and challenges of chatbots in education, it was stated that most of the existing research used pre-test-post-test and questionnaires to analyze the effectiveness of chatbots in education.

Conclusion

The advantages of AI chatbots in increasing student performance and motivation, decreasing instructor workload and improving educational resources show that this tool is an instructional technology tool that can be used to

solve learning-related problems. It is necessary to determine the appropriate conditions and methods for the use of chatbots in education. In the planning and execution of learning activities, the teaching staff's orientation towards activities that will emphasize the advantages and eliminate the disadvantages instead of traditional teaching will help AI-supported chatbots to transform education. The disadvantages associated with AI chatbots can also provide advantages for educational institutions. For example, while plagiarized assignments can be detected, teaching staff will now be able to turn to evaluation methods that will not be stolen from somewhere. This change in assessment processes can have a positive impact on all levels of education.

Language and people are at the forefront of research on the use of AI chatbots in education. It is predicted that the number of studies will increase exponentially. More research is needed to explore the full potential of AI chatbots in education and identify best practices for their implementation. In future research, it is recommended to conduct research on the effects of AI chatbots on pedagogical practices, technology acceptance, usability, individual characteristics, and instructional design processes. Future research is recommended to investigate AI chatbots in the context of online learning communities (Gökçearslan, 2013) and research communities. Although ChatGPT seems to be at the forefront among AI chatbots compared to other tools, different features of hundreds of AI chatbots can be used in different learning/teaching activities. Quantitative method was frequently preferred, while quasi-experimental design and systematic review in qualitative research were frequently preferred. Mixed research method is the least studied research method. In the statistics used in the studies, tests such as frequency percentage and t- test are at the forefront. It is recommended that future studies should be conducted using advanced statistics. About 2/5 of the publications were conducted with undergraduate and graduate groups. Research can be conducted with other study groups, especially with adult and educator groups. About 2 out of 5 publications were conducted with groups of 31-100 people. Self-report data collection tools were frequently used in the studies. It is recommended to use alternative measurement tools in research.

Recommendations

This study has some limitations. In the research, only studies published in journals in the SSCI index were analyzed. There are also research articles on the use of AI chatbots for educational purposes in other databases such as ERIC and ProQuest or in national/international resources. In this systematic review, the document type is "article". Full papers, books, theses, dissertations, and review articles can also be examined.

References

Adamopoulou, E., & Moussiades, L. (2020). Chatbots: History, technology, and applications. *Machine Learning with Applications*, *2*, 100006. 10.1016/j.mlwa.2020.100006

Akçayır, M., & Akçayır, G. (2017). Advantages and challenges associated with augmented reality for education: A systematic review of the literature. *Educational Research Review, 20,* 1-11. 10.1016/j.edurev.2016.11.002

Baidoo-Anu, D., & Owusu Ansah, L. (2023). Education in the era of generative artificial intelligence (AI): Understanding the potential benefits of ChatGPT in promoting teaching and learning. Available at SSRN

4337484. 10.2139/ssrn.4337484

Banihashem, S. K., Dehganzadeh, H., Clark, D., Noroozi, O., & Biemans, H. J. A. (2023). Learning analytics for online game-based learning: A systematic literature review. *Behaviour & Information Technology*, 1-28. https://doi.org/10.1080/0144929X.2023.2255301.

Bozkurt, A. (2023). Generative artificial intelligence (AI) powered conversational educational agents: The inevitable paradigm shift. *Asian Journal of Distance Education*, *18*(1). Retrieved from http://www.asianjde.com/ojs/index.php/AsianJDE/article/view/718

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., & Amodei, D. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, *33*, 1877-1901.

Buchholz, O. (2023). A means-end account of explainable artificial intelligence. *Synthese*, *202*(2), 1-33. https://doi.org/10.1007/s11229-023-04260-w

Chang, C. Y., Hwang, G. J., & Gau, M. L. (2022). Promoting students' learning achievement and self-efficacy: A mobile chatbot approach for nursing training. *British Journal of Educational Technology*, *53*(1), 171-188. https://doi.org/10.1111/bjet.13158

Chen, L., Chen, P., & Lin, Z. (2020). Artificial intelligence in education: A review. *IEEE Access*, *8*, 75264-75278. 10.1109/access.2020.2988510

Chen, X., Cheng, G., Zou, D., Zhong, B., & Xie, H. (2023). Artificial intelligent robots for precision education: A topic modeling based bibliometric analysis. *Educational Technology and Society*, *26*(1), 171-186. https://doi.org/10.30191/ETS.202301_26(1).0013

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

Cotton, D. R., Cotton, P. A., & Shipway, J. R. (2023). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International*, 1-12. https://doi.org/10.1080/14703297.2023.2190148

Dehghanzadeh, H., Farrokhnia, M., Dehghanzadeh, H., Taghipour, K., & Noroozi, O. (2023). Using gamification to support learning in K-12 education: A systematic literature review. *British Journal of Educational Technology*, 00, 1-37. https://doi.org/10.1111/bjet.13335

Delgado-von-Eitzen, C., Anido-Rifón, L., & Fernández-Iglesias, M. J. (2021). Blockchain applications in education: A systematic literature review. *Applied Sciences*, *11*(24), 11811. 10.3390/app112411811

Deveci Topal, A., Dilek Eren, C., & Kolburan Geçer, A. (2021). Chatbot application in a 5th grade science course. *Education and Information Technologies*, *26*(5), 6241-6265. https://doi.org/10.1007/s10639-021-10627-8

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *Arxiv Preprint, arXiv:1810.04805*. 10.48550/arXiv.1810.04805

Dogan, M. E., Goru Dogan, T., & Bozkurt, A. (2023). The use of artificial intelligence (AI) in online learning and distance education processes: A systematic review of empirical studies. *Applied Sciences*, *13*(5), 3056. 10.3390/app13053056

Durall, E., & Kapros, E. (2020). Co-design for a competency self-assessment chatbot and survey in science education. In *Learning and Collaboration Technologies. Human and Technology Ecosystems*, *7th*

International Conference, LCT 2020, Held as Part of the 22nd HCI International Conference, HCII 2020, Copenhagen, Denmark, July 19–24, 2020, Proceedings, Part II 22 (pp. 13-24). Springer International Publishing. https://doi.org/10.1007/978-3-030-50506-6_2

Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., ... & Wright, R. (2023). So what if ChatGPT wrote it?: Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management*, *71*, 102642. 10.1016/j.ijinfomgt.2023.102642

Educause (2023). 2023 Educause horizon report: Teahing and learning edition. https://library.educause.edu/resources/2023/5/2023-educause-horizon-report-teaching-and-learning-edition

Essel, H. B., Vlachopoulos, D., Tachie-Menson, A., Johnson, E. E., & Baah, P. K. (2022). The impact of a virtual teaching assistant (chatbot) on students' learning in Ghanaian higher education. *International Journal of Educational Technology in Higher Education*, *19*(1), 57. https://doi.org/10.1186/s41239-022-00362-6

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*, 1-15. https://doi.org/10.1080/14703297.2023.2195846

Godwin-Jones, R. (2023). Emerging spaces for language learning: AI bots, ambient intelligence, and the metaverse. *Language Learning & Technology*, *27*(2), 6–27. https://hdl.handle.net/10125/73501

González-Calatayud, V., Prendes-Espinosa, P., & Roig-Vila, R. (2021). Artificial intelligence for student assessment: A systematic review. *Applied Sciences*, *11*(12), 5467. https://doi.org/10.3390/app11125467

Goswami, R. (2023). OpenAI changed its plans and won't train on customer data, Sam Altman says. https://www.cnbc.com/2023/05/05/sam-altman-openai-wont-tap-into-customerapis.html?utm_campaign=DonanimHaber&utm_medium=referral&utm_source=DonanimHaber

Gökçearslan, Ş. (2013). Çevrimiçi öğrenme topluluğu hissi ölçeğinin geliştirilmesi: Geçerlik ve güvenirlik çalışması. *Türk Kütüphaneciliği, 27*(1), 154-165.

Guo, K., Wang, J., & Chu, S. K. W. (2022). Using chatbots to scaffold EFL students' argumentative writing. *Assessing Writing*, *54*, 100666. https://doi.org/10.1016/j.asw.2022.100666

Gursoy, D., & Sandstrom, J. K. (2016). An updated ranking of hospitality and tourism journals. *Journal of Hospitality & Tourism Research*, *40*(1), 3-18. 10.1177/1096348014538054

Han, J. W., Park, J., & Lee, H. (2022). Analysis of the effect of an artificial intelligence chatbot educational program on non-face-to-face classes: a quasi-experimental study. *BMC Medical Education*, *22*(1), 830. https://doi.org/10.1186/s12909-022-03898-3

Hsu, T. C., Huang, H. L., Hwang, G. J., & Chen, M. S. (2023). Effects of incorporating an expert decision-making mechanism into chatbots on students' achievement, enjoyment, and anxiety. *Educational Technology & Society*, *26*(1), 218-231. https://www.jstor.org/stable/48707978

Hsu, Y. C., Hung, J. L., & Ching, Y. H. (2013). Trends of educational technology research: More than a decade of international research in six SSCI-indexed refereed journals. *Educational Technology Research and Development*, *61*, 685-705.10.1007/s11423-013-9290-9

Hwang, G. J., & Chang, C. Y. (2021). A review of opportunities and challenges of chatbots in education. *Interactive Learning Environments. 10*.1080/10494820.2021.1952615

Hwang, G. J., & Tu, Y. F. (2021). Roles and research trends of artificial intelligence in mathematics education: A bibliometric mapping analysis and systematic review. *Mathematics*, *9*(6), 584. https://doi.org/10.3390/math9060584

Ji, H., Han, I., & Ko, Y. (2023). A systematic review of conversational AI in language education: Focusing on the collaboration with human teachers. *Journal of Research on Technology in Education*, *55*(1), 48-63. https://doi.org/10.1080/15391523.2022.2142873

Kassab, M., DeFranco, J., & Laplante, P. (2020). A systematic literature review on Internet of things in education: Benefits and challenges. *Journal of Computer Assisted Learning*, *36*(2), 115-127. 10.1111/jcal.12383

Kavanagh, S., Luxton-Reilly, A., Wuensche, B., & Plimmer, B. (2017). A systematic review of virtual reality in education. *Themes in Science and Technology Education*, *10*(2), 85-119.

Kuhail, M. A., Alturki, N., Alramlawi, S., & Alhejori, K. (2023). Interacting with educational chatbots: A systematic review. *Education and Information Technologies*, *28*(1), 973-1018. https://doi.org/10.1007/s10639-022-11177-3

Liang, J. C., Hwang, G. J., Chen, M. R. A., & Darmawansah, D. (2023). Roles and research foci of artificial intelligence in language education: an integrated bibliographic analysis and systematic review approach. *Interactive Learning Environments*, *31*(7), 4270-4296. https://doi.org/10.1080/10494820.2021.1958348

Liu, C. C., Liao, M. G., Chang, C. H., & Lin, H. M. (2022). An analysis of children' interaction with an AI chatbot and its impact on their interest in reading. *Computers & Education*, *189*, 104576. https://doi.org/10.1016/j.compedu.2022.104576

Loafers (2023). Chat GPT 4 release date: OpenAI to launch latest version of AI chatbot next week. https://opchatgptai.com/chat-gpt-4-release-date/

Luor, T., Johanson, R. E., Lu, H. P., & Wu, L. L. (2008). Trends and lacunae for future computer assisted learning (CAL) research: An assessment of the literature in SSCI journals from 1998-2006. *Journal of the American Society for Information Science and Technology*, *59*(8), 1313-1320.10.1002/asi.20836

Mhlanga, D. (2023). Open AI in education, the responsible and ethical use of ChatGPT towards lifelong learning. *Education, the Responsible and Ethical Use of ChatGPT Towards Lifelong Learning (February 11, 2023)*. http://dx.doi.org/10.2139/ssrn.4354422

Miles, M.B. & Huberman, A.M. (1994). Qualitative data analysis. California: SAGE Publications Inc.

Moldt, J. A., Festl-Wietek, T., Madany Mamlouk, A., Nieselt, K., Fuhl, W., & Herrmann-Werner, A. (2023). Chatbots for future docs: exploring medical students' attitudes and knowledge towards artificial intelligence and medical chatbots. *Medical Education Online*, *28*(1). https://doi.org/10.1080/10872981.2023.2182659.

Nee, C. K., Rahman, M. H. A., Yahaya, N., Ibrahim, N. H., Razak, R. A., & Sugino, C. (2023). Exploring the Trend and Potential Distribution of Chatbot in Education: A Systematic Review. *International Journal of Information and Education Technology*, *13*(3), 516-525. 10.18178/ijiet.2023.13.3.1834

Okonkwo, C. W., & Ade-Ibijola, A. (2021). Chatbots applications in education: A systematic review. *Computers and Education: Artificial Intelligence*, *2*, 100033. 10.1016/j.caeai.2021.100033

OpenAI (2023). Optimizing language models for dialogue-OpenAI. Available online:

https://openai.com/blog/chatgpt/

Paek, S., & Kim, N. (2021). Analysis of worldwide research trends on the ımpact of artificial ıntelligence in education. *Sustainability, 13*(14), 7941. MDPI AG. http://dx.doi.org/10.3390/su13147941

Pérez, J. Q., Daradoumis, T., & Puig, J. M. M. (2020). Rediscovering the use of chatbots in education: A systematic literature review. *Computer Applications in Engineering Education*, *28*(6), 1549-1565. 10.1002/cae.22326

Rose, R., Holmes, W., Griffiths, M., Forcier, L. B. (2016). *Intelligence unleashed.* An Argument for AI in Education, Pearson: London, UK.

Rospigliosi, P. A. (2023). Artificial intelligence in teaching and learning: what questions should we ask of ChatGPT?. *Interactive Learning Environments, 31*(1), 1-3. 10.1080/10494820.2023.2180191

Sabzalieva, E., & Valentini, A. (2023). ChatGPT and artificial intelligence in higher education: Quick start guide. https://eduq.info/xmlui/handle/11515/38828

Salas-Pilco, S. Z., Xiao, K., & Hu, X. (2022). Artificial intelligence and learning analytics in teacher education: A systematic review. *Education Sciences*, *12*(8), 569. https://doi.org/10.3390/educsci12080569

Sapçı, A. H., & Sapçı, H. A. (2020). Artificial intelligence education and tools for medical and health informatics students: systematic review. *JMIR Medical Education*, *6*(1), e19285. https://doi.org/10.2196/19285

Smutny, P., & Schreiberova, P. (2020). Chatbots for learning: A review of educational chatbots for the Facebook Messenger. *Computers & Education*, *151*, 103862. 10.1016/j.compedu.2020.103862

Sozbilir, M., Kutu, H., Yasar, M.D. (2012). Science Education Research in Turkey. In: Jorde, D., Dillon, J. (eds) Science Education Research and Practice in Europe. *Cultural Perpectives in Science Education*, *5*. SensePublishers, Rotterdam. https://doi.org/10.1007/978-94-6091-900-8_14

Tang, K. Y., Chang, C. Y., & Hwang, G. J. (2023). Trends in artificial intelligence-supported e-learning: A systematic review and co-citation network analysis (1998-2019). *Interactive Learning Environments*, *31*(4), 2134-2152. https://doi.org/10.1080/10494820.2021.1875001

Tesch, R. (2013). *Qualitative research: Analysis types and software*. Routledge. 10.4324/9781315067339

UNESCO (2022). K-12 AI curricula: A mapping of government-endorsed AI curricula. https://unesdoc.unesco.org/ark:/48223/pf0000380602

Vázquez-Cano, E., Mengual-Andrés, S., & López-Meneses, E. (2021). Chatbot to improve learning punctuation in Spanish and to enhance open and flexible learning environments. *International Journal of Educational Technology in Higher Education*, *18*(1), 1-20. https://doi.org/10.1186/s41239-021-00269-8

Wollny, S., Schneider, J., Di Mitri, D., Weidlich, J., Rittberger, M., & Drachsler, H. (2021). Are we there yet?-a systematic literature review on chatbots in education. *Frontiers in Artificial Intelligence*, *4*, 654924. https://doi.org/10.3389/frai.2021.654924

Xia, Q., Chiu, T. K., Chai, C. S., & Xie, K. (2023). The mediating effects of needs satisfaction on the relationships between prior knowledge and self-regulated learning through artificial intelligence chatbot. *British Journal of Educational Technology*. https://doi.org/10.1111/bjet.13305

Xu, Y., Wang, D., Collins, P., Lee, H., & Warschauer, M. (2021). Same benefits, different communication patterns: Comparing Children's reading with a conversational agent vs. a human partner. *Computers & Education*, *161*, 104059. https://doi.org/10.1016/j.compedu.2020.104059

Yang, H., Kim, H., Lee, J., & Shin, D. (2022). Implementation of an AI chatbot as an English conversation partner

in EFL speaking classes. *ReCALL, 34*(3), 327-343. https://doi.org/10.1017/S0958344022000039

Yang, Z., Dai, Z., Yang, Y., Carbonell, J., Salakhutdinov, R. R., & Le, Q. V. (2019). Xlnet: Generalized autoregressive pretraining for language understanding. *Advances in Neural Information Processing Systems* (Edited by: H. Wallach and H. Larochelle and A. Beygelzimer and F. d'Alché-Buc and E. Fox and R. Garnett), *32*.

Yiğitcanlar, T., Desouza, K. C., Butler, L., & Roozkhosh, F. (2020). Contributions and risks of artificial intelligence (AI) in building smarter cities: Insights from a systematic review of the literature. *Energies*, *13*(6), 1473. https://doi.org/10.3390/en13061473

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education–where are the educators?. *International Journal of Educational Technology in Higher Education, 16*(1), 1-27. https://doi.org/10.1186/s41239-019-0171-0

Zhang, K., & Aslan, A. B. (2021). AI technologies for education: Recent research & future directions. Computers and Education: Artificial Intelligence, 2, 100025. 10.48550/arXiv.1906.08237

Zhang, M., & Li, J. (2021). A commentary of GPT-3 in MIT technology review 2021. *Fundamental Research*, *1*(6), 831-833. 10.1016/j.fmre.2021.11.011

Author Information

Şahin Gökçearslan

http://orcid.org/0000-0003-3532-4251

Gazi University

Dıstance Educatıon R & D Center

Turkiye

Contact e-mail: *sahingokce@gmail.com*

Cansel Tosun

https://orcid.org/0009-0009-2461-6508

Gazi University

Gazi Education Faculty

Turkiye

Zeynep Gizem Erdemir

https://orcid.org/0009-0002-3696-8411

Gazi University

Computer Education and Instructional Technology

Gazi Education Faculty

Turkiye